

nature

THE BITER BIT
Viral infections for viruses

TROPICAL CYCLONES
The strong get stronger

BLACK HOLE PHYSICS
A new window on the
Galactic Centre



**BIG
DATA**

NATURE.COM
Microarray technology

**SCIENCE IN THE
PETABYTE ERA**

Community cleverness required

Researchers need to adapt their institutions and practices in response to torrents of new data — and need to complement smart science with smart searching.

The Internet search firm Google was incorporated just 10 years ago this week. Going from a collection of donated servers housed under a desk to a global network of dedicated data centres processing information by the petabyte, Google's growth mirrors that of the production and exploration of data in research. All of which makes this an apt moment for this special issue of *Nature*, which examines what big data sets mean for contemporary science.

'Big', of course, is a moving target. The portability of the tens of gigabytes we carry around on USB sticks would have seemed like fantasy a few years ago. But beyond a certain point, as an increasing number of research disciplines are discovering, the vast amounts of data are presenting fresh challenges that urgently need to be addressed.

The issue is partly a matter of the sheer scale of today's data sets. Managing this torrent of bits has forced more and more fields to move to industrial-scale data centres and cutting-edge networking technology (see page 16). But the data sets are also becoming increasingly complex. As researchers study the inner workings of the cell, for example, they now gather data on genomic sequences, protein sequences, protein structure and function, bimolecular interactions, signalling and metabolic pathways, regulatory motifs — on and on. No wonder even the smartest scientists turn with relief to advanced data-mining tools, online community collaborations (see page 22) and sophisticated visualization techniques (see page 30).

Sudden influxes of data have transformed researchers' understanding of nature before — even back in the days when 'computer' was still a job description (see page 36). Unfortunately, the institutions and culture of science remain rooted in that pre-electronic era. Taking full advantage of electronic data will require a great deal of additional infrastructure, both technical and cultural (see pages 8, 28 and 47).

The lack of standards, for instance, confounds many a researcher seeking to harness the diversity of knowledge now available on any chosen topic. All

credit, then, to those in the vanguard of interoperability. In biology, for example, the Gene Ontology Consortium has spent the past decade devising consistent descriptions of gene products in different databases. Meanwhile, the Mouse Genome Informatics resource is a good demonstrator of complexity's challenges and solutions. Funding agencies have been slow to support data infrastructure and this is one cultural shift that needs to accelerate — although recent efforts by the US National Science Foundation and Germany's DFG are a good beginning. But above all, such standards require support from researchers, who should adopt them and deploy them consistently. This takes a degree of intellectual and practical commitment to what can seem like tedious bookkeeping.

Researchers need to be obliged to document and manage their data with as much professionalism as they devote to their experiments. And they should receive greater support in this endeavour than they are afforded at present. Those publicly funded databases that have taken on preservation responsibilities, such as GenBank and UniProt, are only a small part of the data landscape. Universities and funding agencies need to provide and support curation facilities, tools and training.

As is amply highlighted in this issue, all of these worthy aims require incentives. These include pressure from, and recognition through, journals. *Nature* and its sister publications have always worked closely with those developing databases and standards, and we remain committed to continuing such community collaborations. Incentives also include recognition of impactful informatics by peer committees and research-rating exercises.

Above all, data on today's scales require scientific and computational intelligence. Google may now have its critics, but no one can deny its impact, which ultimately stems from the cleverness of its informatics. The future of science depends in part on such cleverness again being applied to data for their own sake, complementing scientific hypotheses as a basis for exploring today's information cornucopia. ■

EDITORIAL

- 1 **Community cleverness required**



NEWS

- 8 **SPECIAL REPORT The next Google**
Duncan Graham-Rowe

PARTY OF ONE

- 15 **Data wrangling**
David Goldston

NEWS FEATURES

- 16 **Welcome to the petacentre**
Cory Doctorow
- 22 **Wikionomics**
Mitch Waldrop

COMMENTARY

- 28 **How do your data grow?**
Clifford Lynch

BOOKS & ARTS

- 30 **Distilling meaning from data**
Felice Frankel & Rosalind Reid

ESSAY

- 36 **The Harvard computers**
Sue Nelson

FEATURE

- 47 **The future of biocuration**
Doug Howe, Seung Yon Rhee *et al.*



For podcast and more online extras see www.nature.com/news/specials/bigdata/

Cool philosophies

High-energy physicists should not gloss over fundamental conundrums.

Whether through calculated seduction or natural attraction, the Large Hadron Collider (LHC), which starts up next week, seems to be enjoying a love affair with the media and the public. Some may revel in the Herculean feat of engineering at CERN, the particle-physics laboratory near Geneva, that has created a 27-kilometre tunnel colder than the outer cosmos. But the LHC's most potent allure surely stems from the air of mystery that attaches to its promise to reveal secrets from the first milliseconds of the Big Bang — in lay terms, to explain how it all began.

Needless to say, the LHC cannot do quite that. But it should tell us something about the origins of mass, courtesy of the much-vaunted Higgs boson, as well as casting light on the asymmetry between matter and antimatter, and the nature of the quark–gluon plasma that seethed before the Universe was cool enough to make nuclear particles. We might even catch glimpses of the truly exotic: spawnings of tiny black holes and their evaporation, and signs of extra dimensions. The excitement and anticipation are warranted.

All this places a fearsome burden of explanation on the physicists involved. Can they convey anything beyond the sound bites above and, if so, how? As Alexei Grinbaum, a physicist and ethicist at CEA-Saclay in Gif-sur-Yvette, France, writes in a stimulating preprint (www.arxiv.org/abs/0806.4268): “The LHC is an opportunity to renew the enthusiasm for *understanding the world* ... Whether such a change will occur depends ... on how physicists will speak about the LHC and what they will say.” He expresses a hope that the discourse

will not amount to *ex cathedra* pronouncements of the sort: “I can't explain without the full mathematics what is *really* going on.”

But as Grinbaum argues, this is not simply a matter of finding the right, non-alienating tone. When physicists, struggling to put across a difficult concept to a lay audience, say (or more probably just think): “Oh, if only I could show you the equations, you would understand,” this is not what they really mean. Rather, they mean: “You would understand at the same level as I understand.” That is, at the level of mathematical formalism. This is not to imply that physicists hide blindly behind the maths (although some probably do), but that they might not acknowledge or even recognize that the mathematics shields them from genuine conceptual questions.

There is a tendency to wave these questions away as semantic or philosophical, as though such issues by definition cannot be serious. The founders of quantum theory knew otherwise, although some of their dilemmas have still not been resolved. As the philosopher Moritz Schlick said: “It is the mark of the greatest scientific minds that they think out every question they take up right to the end, and the end of every question lies in philosophy.” Grinbaum puts it another way: “The job of [the] theoretical physicist is not to write equations.”

For the LHC, some of these foundational issues are raised by the role of aesthetics as a guide to physical theory, in particular arguments based on symmetry. At the pragmatic level, symmetry has been an immensely fertile tool, and it underpins the notion of a Higgs mechanism for mass. But there is no rigorous justification for relying on it, and it is possible that the LHC might point the way to a new physics that discards it as a ruling principle.

These questions, which impose themselves only when a complacent reliance on ‘the equations’ is renounced, could captivate a wide audience as readily as tantalizing references to the first instants of creation. It would be a shame if the teams at the LHC shy away from them. ■

The hour of diplomacy

Scientific collaboration between East and West must survive the crisis in Georgia.

Less than three months ago, hundreds of scientists from Russia, Europe, Asia and North America gathered in Tbilisi, Georgia, for the 2008 Phage Biology, Ecology and Therapy Meeting at the George Eliava Institute of Bacteriophage, Microbiology and Virology. In the light of recent events, it seems unlikely that a similarly illustrious scientific meeting will happen again there any time soon.

But the political reverberations that have inevitably followed the conflict in the Caucasus region should not be allowed to damage science. Indeed, the rougher the language between Moscow and the West, the more valuable it becomes to sustain good relations in research.

Science cannot and should not be blind to politics. But even if the current political crisis escalates, it would be utterly unwise of either side to halt or suspend existing scientific agreements and collaborations. Over the past two decades, East–West scientific relationships have developed from pure unilateral aid to increasingly fruitful collaborations.

This process must continue. Joint efforts in arms control, nuclear

non-proliferation and biosafety have made this planet a safer place to live than it was at the height of the cold war. Mechanisms such as the International Science and Technology Center, which since 1992 has provided former weapons scientists with opportunities in international partnerships, have helped prevent Soviet research centres and armouries from becoming shopping paradises for terrorists and dodgy heads-of-state. And the influx to Western labs of countless highly skilled Russian mathematicians, computer scientists and others has propelled advances in fields ranging from bioinformatics to climate modelling.

Looking forward, collaboration with Russia in space, energy and arms limitation will be a key issue for the incoming US administration, and the European Union's member states need to decide whether Russia can join the seventh Framework programme (see page 6).

The current political tensions could all too easily block these avenues of collaboration, which would harm not only Western interests but those of Russia's researchers too. Scientists and political leaders should do everything they can to avoid those outcomes. As Russia regains its strength and pride, direct financial support for Russian science from foreign governments will inevitably decline. What should follow is an equal scientific partnership based on mutual trust, respect and responsibility. Prudent scientific diplomacy is a peace-keeping measure in its own right. ■

RESEARCH HIGHLIGHTS

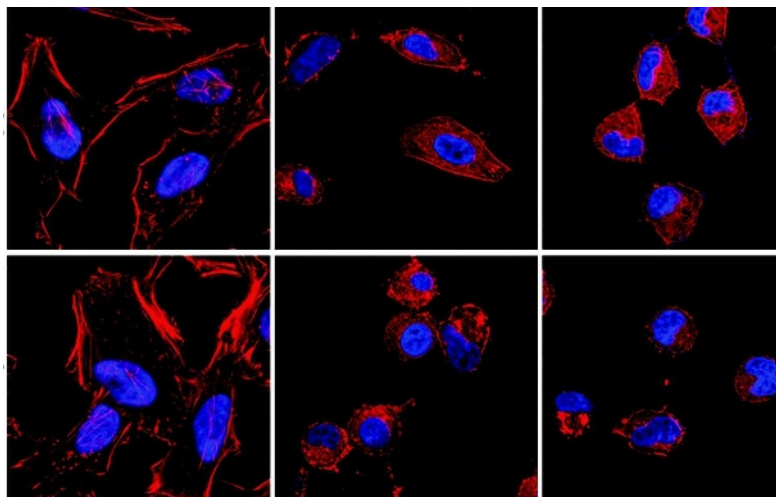
Cause of death

Proc. Natl Acad. Sci. USA **105**, 12497–12502 (2008)

The bacterial pathogen *Vibrio parahaemolyticus* has an unusual way of killing cells, researchers report.

V. parahaemolyticus causes severe diarrhoea and can be life threatening in people with weakened immune systems. Previous work had suggested that the bacteria kill cells by injecting them with proteins that trigger a kind of cellular suicide called apoptosis. But Kim Orth and her colleagues at the University of Texas Southwestern Medical Center have found that dying cells do not express enzymes characteristic of this.

Instead, the cells were inflamed. They became more rounded (progression pictured left to right for two samples), began to degrade their own contents and started to leak. These events occurred in only a few hours, and together provide a new means by which a pathogen that works from outside a cell can kill its target.



D.L. BURDETTE ET AL./PNAS

GEOSCIENCES**Clubmoss clues**

Nature Geosci. doi:10.1038/ngeo278 (2008)

Spores in herbaria around the world may help to push back the record of atmospheric ozone concentrations, according to Barry Lomax, now at the University of Nottingham, UK, and his co-workers.

They think that the quantities of two ultraviolet-B-absorbing compounds (*p*-coumaric and ferulic acid) present in the outer walls of spores and pollen can serve as a proxy for stratospheric ozone. This is because the less ozone there is in the atmosphere, the more UV-B radiation reaches Earth, and the more UV-B-protecting chemicals plants make.

Lomax's team tested the idea on two species of clubmoss: *Lycopodium magellanicum* and *L. annotinum*, from which they reconstructed UV-B flux back to 1907 — 20 years earlier than any previous record. Spores from Greenland and South Georgia, an island in the South Atlantic, gave the same pattern.

PHYSICS**A bolt from the blue**

Phys. Res. Lett. **101**, 075005 (2008)

Bolts of lightning expand by sending out trees of tiny filaments called streamers, which are ionized air channels. Logic dictates that the channel heads should repel one another because they carry the same electric charge. But, as any lightning-watcher can vouch, streamers touch quite often (pictured right).

Ute Ebert and her colleagues at the National Research Institute for Mathematics and Computer Science in Amsterdam, The Netherlands, have simulated how streamers connect. They have demonstrated that, in a gaseous mixture of nitrogen and oxygen

(as in Earth's atmosphere), streamers come together more easily at lower pressures. Thus their model could explain the observations. The work paves the way to simulating a complete lightning fork, not just single streamers.

MOLECULAR BIOLOGY**Precision dumping**

Cell **134**, 668–678 (2008)

Editing a molecular tag called polyubiquitin sends two key immune-response proteins into the cellular garbage-disposal system.

There are two main ways of attaching one ubiquitin to another within polyubiquitin, and the one chosen often determines whether the target protein is activated or degraded. Vishva Dixit at Genentech in South San Francisco, California, and his team have made antibodies that can discriminate between the two.

Using these antibodies, they discovered that RIP1 and IRAK1 — proteins involved in a cell's response to immune-system signals — start

out with the activating type of attachment, and that this is later 'edited' into the degrading one. This editing could be a way of dampening down other cell-signalling pathways.

CHEMISTRY**Silicon pulls it off**

Science **321**, 118–190 (2008)

Organic molecules containing carbon–fluorine bonds are long-lived atmospheric pollutants that act as powerful greenhouse gases. The secret of their longevity is the stubbornly unreactive nature of these bonds. A catalyst that could turn those tough C–F bonds into C–H bonds quickly and selectively would be a boon to people disposing of ozone-unfriendly molecules of this sort. Christos Douvris and Oleg Ozerov of Brandeis University in Waltham, Massachusetts, have designed one.

Their catalyst contains silicon, which facilitates a bond-swapping reaction. Because the C–F bond is weaker than the Si–F bond and the Si–H bond weaker than the C–H bond, the switch from C–F to C–H by way of Si is thermodynamically favourable. The reaction occurs in mild conditions and the catalyst is reusable.

IMMUNOLOGY**Hitting 'pause'**

Cell **134**, 657–667 (2008)

Although senescent cells may appear dormant, they have an important role in protecting the liver from cirrhosis, according to Scott Lowe of Cold Spring Harbor Laboratory in New York and his colleagues.

Cirrhosis results from the long-term accumulation of fibrous scar tissue. Lowe and his team discovered that senescent cells



R. WETMORE/GETTY

RESEARCH HIGHLIGHTS

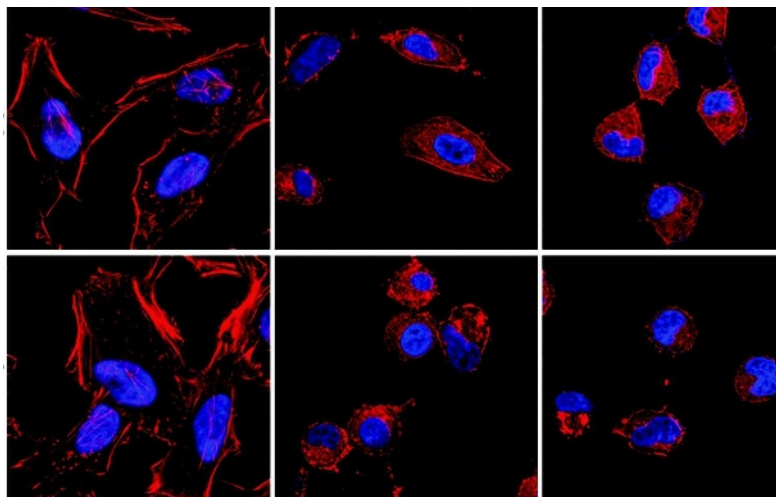
Cause of death

Proc. Natl Acad. Sci. USA **105**, 12497–12502 (2008)

The bacterial pathogen *Vibrio parahaemolyticus* has an unusual way of killing cells, researchers report.

V. parahaemolyticus causes severe diarrhoea and can be life threatening in people with weakened immune systems. Previous work had suggested that the bacteria kill cells by injecting them with proteins that trigger a kind of cellular suicide called apoptosis. But Kim Orth and her colleagues at the University of Texas Southwestern Medical Center have found that dying cells do not express enzymes characteristic of this.

Instead, the cells were inflamed. They became more rounded (progression pictured left to right for two samples), began to degrade their own contents and started to leak. These events occurred in only a few hours, and together provide a new means by which a pathogen that works from outside a cell can kill its target.



D.L. BURDETTE ET AL./PNAS

GEOSCIENCES**Clubmoss clues**

Nature Geosci. doi:10.1038/ngeo278 (2008)

Spores in herbaria around the world may help to push back the record of atmospheric ozone concentrations, according to Barry Lomax, now at the University of Nottingham, UK, and his co-workers.

They think that the quantities of two ultraviolet-B-absorbing compounds (*p*-coumaric and ferulic acid) present in the outer walls of spores and pollen can serve as a proxy for stratospheric ozone. This is because the less ozone there is in the atmosphere, the more UV-B radiation reaches Earth, and the more UV-B-protecting chemicals plants make.

Lomax's team tested the idea on two species of clubmoss: *Lycopodium magellanicum* and *L. annotinum*, from which they reconstructed UV-B flux back to 1907 — 20 years earlier than any previous record. Spores from Greenland and South Georgia, an island in the South Atlantic, gave the same pattern.

PHYSICS**A bolt from the blue**

Phys. Res. Lett. **101**, 075005 (2008)

Bolts of lightning expand by sending out trees of tiny filaments called streamers, which are ionized air channels. Logic dictates that the channel heads should repel one another because they carry the same electric charge. But, as any lightning-watcher can vouch, streamers touch quite often (pictured right).

Ute Ebert and her colleagues at the National Research Institute for Mathematics and Computer Science in Amsterdam, The Netherlands, have simulated how streamers connect. They have demonstrated that, in a gaseous mixture of nitrogen and oxygen

(as in Earth's atmosphere), streamers come together more easily at lower pressures. Thus their model could explain the observations. The work paves the way to simulating a complete lightning fork, not just single streamers.

MOLECULAR BIOLOGY**Precision dumping**

Cell **134**, 668–678 (2008)

Editing a molecular tag called polyubiquitin sends two key immune-response proteins into the cellular garbage-disposal system.

There are two main ways of attaching one ubiquitin to another within polyubiquitin, and the one chosen often determines whether the target protein is activated or degraded. Vishva Dixit at Genentech in South San Francisco, California, and his team have made antibodies that can discriminate between the two.

Using these antibodies, they discovered that RIP1 and IRAK1 — proteins involved in a cell's response to immune-system signals — start

out with the activating type of attachment, and that this is later 'edited' into the degrading one. This editing could be a way of dampening down other cell-signalling pathways.

CHEMISTRY**Silicon pulls it off**

Science **321**, 118–190 (2008)

Organic molecules containing carbon–fluorine bonds are long-lived atmospheric pollutants that act as powerful greenhouse gases. The secret of their longevity is the stubbornly unreactive nature of these bonds. A catalyst that could turn those tough C–F bonds into C–H bonds quickly and selectively would be a boon to people disposing of ozone-unfriendly molecules of this sort. Christos Douvris and Oleg Ozerov of Brandeis University in Waltham, Massachusetts, have designed one.

Their catalyst contains silicon, which facilitates a bond-swapping reaction. Because the C–F bond is weaker than the Si–F bond and the Si–H bond weaker than the C–H bond, the switch from C–F to C–H by way of Si is thermodynamically favourable. The reaction occurs in mild conditions and the catalyst is reusable.

IMMUNOLOGY**Hitting 'pause'**

Cell **134**, 657–667 (2008)

Although senescent cells may appear dormant, they have an important role in protecting the liver from cirrhosis, according to Scott Lowe of Cold Spring Harbor Laboratory in New York and his colleagues.

Cirrhosis results from the long-term accumulation of fibrous scar tissue. Lowe and his team discovered that senescent cells



R. WETMORE/GETTY

accumulate in cirrhotic livers and form the structural basis for fibrosis. The researchers then gave a liver-damaging compound to mutant mice that lacked a protein required for senescence. These animals suffered more liver fibrosis than normal mice.

Other experiments revealed that senescent cells seem to attract the attention of immune cells called natural killer cells. These target the senescent cells for destruction, aiding in a partial recovery from liver fibrosis.

GEOSCIENCES

Goodbye April showers

Geophys. Res. Lett. doi:10.1029/2008GL034828 (2008)

Climate models predict that as Earth warms, the Northern Annular Mode (NAM), a flip-flopping pattern of climate variability in the Northern Hemisphere, will flop more firmly to its low-pressure-near-the-pole state.

By studying climate records, Stephanie McAfee and Joellen Russell of the University of Arizona in Tucson have shown what this means for spring weather in the southwestern United States — spring weather being particularly sensitive to the NAM's behaviour. An intensified NAM leads to warm weather coming earlier, shortening the winter rainy season and leading to drier weather that year. Their findings agree with work on changes in the path of the jet stream, and thus in the tracks that storms follow. Broadly speaking, more storms tracking further to the north mean less rain in the south.

NEUROSCIENCE

Coke heads

Neuron **59**, 621–633 (2008)

Repeated exposure to cocaine increases the density of connections among the nerve cells in a brain region that is central to motivation and reward. The change seems to stymie long-lasting behaviours associated with chronic drug use rather than promote them, as scientists had previously thought.

Working with mice, Christopher Cowan at the University of Texas Southwestern Medical Center in Dallas and his colleagues have found that cocaine suppresses the protein MEF2, encouraging medium-sized spiny neurons in the nucleus accumbens to form more links. Artificially upping the levels of MEF2 blocked this process, as expected, but surprisingly caused mice to behave as though their sensitivity to cocaine had increased.

The researchers propose that MEF2 suppression and the consequent increase in neuronal connections attenuate the harmful effects of long-term cocaine use.

ARCHAEOLOGY

Amazonian urbanites

Science **321**, 1214–1217 (2008)

The Amazon is not a pristine wilderness. In fact, there is increasing evidence for sophisticated town planning there long before Europeans arrived.

Michael Heckenberger at the University of Florida in Gainesville and his band of archaeologists (pictured below) have uncovered a network of settlements around the Xingu River in Brazil. These hamlets were connected by criss-crossing roads that emanated from a central village that was probably more ceremonial than residential.

The dispersed pattern of settlements is unusual. The authors suggest that this arrangement, coupled with the power of Amazonian foliage to overrun abandoned sites, has perhaps blinded researchers to the extent of human impact on the rainforests.



ASTROPHYSICS

Far off fly-by

Astrophys. J. **683**, 722–749 (2008)

M31, the spiral galaxy nearest the Milky Way, and NGC 205, a nearby dwarf elliptical galaxy, appear to be stuck in an eternal *pas de deux*. At least, that is what it looks like through a telescope. But Kirsten Howley of the University of California at Santa Cruz and her colleagues have used what is known as a genetic algorithm to determine that NGC 205 is actually swinging around M31.

The algorithm sifted through more than a billion trillion possible orbits for NGC 205, identifying which of them best fitted the galaxy's observed motions and light characteristics.

Howley's team found that NGC 205 was zipping past M31 at hundreds of kilometres per second, close to its escape velocity. NGC 205's motion is perpendicular to, and therefore independent of, a streamer of stars previously thought to be associated with it.

SCIENCE/AAAS

JOURNAL CLUB

Caroline Harwood

University of Washington, Seattle

A microbiologist hopes to disrupt bacterial 'decisions'.

Cyclic-di-GMP is small but important. It is an intracellular signalling molecule that controls lifestyle choices in bacteria. When should a bacterium become virulent? When should it differentiate into a new cell type? When might it do better to stop moving around and stay still with many others? Bacteria that gather together tend to encase themselves and their neighbours in a carbohydrate slime, forming what is known as a biofilm. I, like many microbiologists, am keen to find ways to disrupt biofilms, and a better understanding of how cyclic-di-GMP works may provide a way to do this.

Recently, answers have started to emerge. First it was shown that cyclic-di-GMP can bind to certain proteins that modulate the activity of flagellar motors — which propel free-swimming bacteria — and to enzymes that make the biofilm-cementing slime. Then researchers found a protein that 'turns on' some of the slime genes when it attaches to cyclic-di-GMP. But one paper shows a completely new way in which cyclic-di-GMP can control bacterial lifestyle choices: by binding to a regulatory region, called a riboswitch, on a messenger RNA molecule (N. Sudarsan *et al.* *Science* **321**, 411–413; 2008).

Ronald Breaker and his team at Yale University in New Haven, Connecticut, report how they used various molecular-biology techniques to demonstrate that part of the RNA hitches itself to cyclic-di-GMP. They also proved that cyclic-di-GMP-binding riboswitches from several bacterial strains can function as genetic 'off' as well as 'on' switches.

These findings are noteworthy because humans do not make cyclic-di-GMP, so the molecule could be a target for new antibiotics. Medicines that attack cyclic-di-GMP should be able to treat biofilm-related disorders such as periodontal disease and ear infections, which are often resistant to existing drugs.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

accumulate in cirrhotic livers and form the structural basis for fibrosis. The researchers then gave a liver-damaging compound to mutant mice that lacked a protein required for senescence. These animals suffered more liver fibrosis than normal mice.

Other experiments revealed that senescent cells seem to attract the attention of immune cells called natural killer cells. These target the senescent cells for destruction, aiding in a partial recovery from liver fibrosis.

GEOSCIENCES

Goodbye April showers

Geophys. Res. Lett. doi:10.1029/2008GL034828 (2008)

Climate models predict that as Earth warms, the Northern Annular Mode (NAM), a flip-flopping pattern of climate variability in the Northern Hemisphere, will flop more firmly to its low-pressure-near-the-pole state.

By studying climate records, Stephanie McAfee and Joellen Russell of the University of Arizona in Tucson have shown what this means for spring weather in the southwestern United States — spring weather being particularly sensitive to the NAM's behaviour. An intensified NAM leads to warm weather coming earlier, shortening the winter rainy season and leading to drier weather that year. Their findings agree with work on changes in the path of the jet stream, and thus in the tracks that storms follow. Broadly speaking, more storms tracking further to the north mean less rain in the south.

NEUROSCIENCE

Coke heads

Neuron **59**, 621–633 (2008)

Repeated exposure to cocaine increases the density of connections among the nerve cells in a brain region that is central to motivation and reward. The change seems to stymie long-lasting behaviours associated with chronic drug use rather than promote them, as scientists had previously thought.

Working with mice, Christopher Cowan at the University of Texas Southwestern Medical Center in Dallas and his colleagues have found that cocaine suppresses the protein MEF2, encouraging medium-sized spiny neurons in the nucleus accumbens to form more links. Artificially upping the levels of MEF2 blocked this process, as expected, but surprisingly caused mice to behave as though their sensitivity to cocaine had increased.

The researchers propose that MEF2 suppression and the consequent increase in neuronal connections attenuate the harmful effects of long-term cocaine use.

ARCHAEOLOGY

Amazonian urbanites

Science **321**, 1214–1217 (2008)

The Amazon is not a pristine wilderness. In fact, there is increasing evidence for sophisticated town planning there long before Europeans arrived.

Michael Heckenberger at the University of Florida in Gainesville and his band of archaeologists (pictured below) have uncovered a network of settlements around the Xingu River in Brazil. These hamlets were connected by criss-crossing roads that emanated from a central village that was probably more ceremonial than residential.

The dispersed pattern of settlements is unusual. The authors suggest that this arrangement, coupled with the power of Amazonian foliage to overrun abandoned sites, has perhaps blinded researchers to the extent of human impact on the rainforests.



ASTROPHYSICS

Far off fly-by

Astrophys. J. **683**, 722–749 (2008)

M31, the spiral galaxy nearest the Milky Way, and NGC 205, a nearby dwarf elliptical galaxy, appear to be stuck in an eternal *pas de deux*. At least, that is what it looks like through a telescope. But Kirsten Howley of the University of California at Santa Cruz and her colleagues have used what is known as a genetic algorithm to determine that NGC 205 is actually swinging around M31.

The algorithm sifted through more than a billion trillion possible orbits for NGC 205, identifying which of them best fitted the galaxy's observed motions and light characteristics.

Howley's team found that NGC 205 was zipping past M31 at hundreds of kilometres per second, close to its escape velocity. NGC 205's motion is perpendicular to, and therefore independent of, a streamer of stars previously thought to be associated with it.

SCIENCE/AAAS

JOURNAL CLUB

Caroline Harwood

University of Washington, Seattle

A microbiologist hopes to disrupt bacterial 'decisions'.

Cyclic-di-GMP is small but important. It is an intracellular signalling molecule that controls lifestyle choices in bacteria. When should a bacterium become virulent? When should it differentiate into a new cell type? When might it do better to stop moving around and stay still with many others? Bacteria that gather together tend to encase themselves and their neighbours in a carbohydrate slime, forming what is known as a biofilm. I, like many microbiologists, am keen to find ways to disrupt biofilms, and a better understanding of how cyclic-di-GMP works may provide a way to do this.

Recently, answers have started to emerge. First it was shown that cyclic-di-GMP can bind to certain proteins that modulate the activity of flagellar motors — which propel free-swimming bacteria — and to enzymes that make the biofilm-cementing slime. Then researchers found a protein that 'turns on' some of the slime genes when it attaches to cyclic-di-GMP. But one paper shows a completely new way in which cyclic-di-GMP can control bacterial lifestyle choices: by binding to a regulatory region, called a riboswitch, on a messenger RNA molecule (N. Sudarsan *et al.* *Science* **321**, 411–413; 2008).

Ronald Breaker and his team at Yale University in New Haven, Connecticut, report how they used various molecular-biology techniques to demonstrate that part of the RNA hitches itself to cyclic-di-GMP. They also proved that cyclic-di-GMP-binding riboswitches from several bacterial strains can function as genetic 'off' as well as 'on' switches.

These findings are noteworthy because humans do not make cyclic-di-GMP, so the molecule could be a target for new antibiotics. Medicines that attack cyclic-di-GMP should be able to treat biofilm-related disorders such as periodontal disease and ear infections, which are often resistant to existing drugs.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

Russia's international research ties under threat

Responses to Russia's military action in Georgia have implications for non-proliferation, space exploration, climate negotiation and the European Union's framework programme.

European Union (EU) officials met in Brussels on 1 September to review relationships with Moscow amid growing friction over Russia's conduct in Georgia. Although no sanctions were agreed, the crisis is threatening hopes that Russian science will soon emerge from its state of isolation.

For now, EU science leaders are saying that international collaborations will continue. "Breaking off our scientific relations with Russia is not something that is currently on the radar," a high-ranking European Commission official told *Nature* before the meeting. "Everything depends on the political development, but we do still hope that our partnerships in science will come undamaged through the current crisis."

The primary science and technology agreement between the EU and Russia is a short document stressing the mutual intent to work together in basic science and its development. Signed in 2000 and renewed for 2004, it is up again for renewal next year. The EC official, who did not want to be identified because decisions would be made by the council, says that the science agreement is likely to be extended.

But growing political tension could at least delay Russia joining the Seventh Framework Programme, the main pan-European research funding tool. Russia is keen to become an associated partner by 2010, the state that Israel, Switzerland and a number of other non-EU countries currently enjoy. This would allow individual scientists, research institutes and companies in Russia to compete in proposals for the €53.2-billion (US\$78 billion) programme that runs from 2007 to 2013.

The Russian science ministry has signalled that it is prepared to make a large financial contribution to the programme. But the commission has not yet received a mandate from EU member states to formally take up negotiations with Russia. And policy analysts suggest that former Eastern Bloc states, such as Poland and the Baltic countries, might now balk at approving such a mandate.

Cooperation between Russia and the United States in space is also under fresh strain. Between 2010, when the US space shuttle is due

to be retired, and 2015, when its replacement is slated to be operational, NASA will have to rely on Russian Soyuz spacecraft to taxi astronauts to and from the International Space Station.

Since the Georgia–Russia conflict began, that imminent gap has become more prominent, says Roald Sagdeev, former director of the Russian Space Research Institute in Moscow and now a physicist at the University of Maryland in College Park. "The United States has been on a collision course with Russia for quite a while," he says. "I think we're going to enter a very difficult period."

US politicians are starting to react. In recent weeks, both leading presidential candidates have signalled support for adding another shuttle flight to the list of remaining launches — although a single flight would do little to shrink the time gap. A NASA authorization bill awaiting Congressional approval also calls for more flights. Democratic contender Barack Obama has said he supports boosting NASA's budget by \$2 billion, which could help close the gap. NASA is now also re-evaluating whether to fly the shuttle after 2010, in part because of a letter sent to the White House last week that was co-signed by John McCain, the Republican contender.

Keeping routes open

More immediately, congressional staff say, Congress may now delay extending a waiver to the 2000 Iran Non-Proliferation Act, which forbids dealings with countries that sell nuclear materials to Iran or North Korea. NASA has asked for the waiver this year so it can renew

the contract with Russia to buy flights on Soyuz vehicles; the contract expires in 2011, and Russia needs a three-year head start for any orders. Sagdeev thinks the

two nations will come to an agreement somehow: "My prediction is that Russians will keep this route open for Americans," he says. "I think NASA will also try to keep it open."

Other non-proliferation initiatives have already suffered. Congressional action on a deal promoting broader cooperation between the United States and Russia on nuclear energy — which had a doubtful future even before the



Russian President Dmitry Medvedev at a missile base in Teikovo, Russia, in May 2008.

Georgian conflict — is now all but impossible. The Bush administration is also reconsidering talks with Russia over the Strategic Arms Reduction Treaty, the rules of which expire next year. Unless they are renewed in some fashion, the 2002 Moscow Treaty promising further reductions will be left without any verification protocols, says Laura Holgate, vice-president of the Nuclear Threat Initiative in Washington DC.

Siegfried Hecker, a former director of Los Alamos National Laboratory in New Mexico, now co-director of the Stanford University Center for International Security and Cooperation in California, says that a political realignment is in order. "Clearly the two countries are going to have to re-establish some sort of an appropriate relationship with each other," he says, "and I do believe that the big questions will not be sorted out until we have a new administration." In the meantime, he says, relations between scientists are likely to continue as they always have, even during the cold war. Hecker is headed to Russia this week to discuss a joint conference on non-proliferation, expected to take place next year.

A meeting next spring in Rome of the national science academies of the G8 countries will also go ahead as planned, says Volker Ter Meulen, president of the Leopoldina, Germany's national academy of science. "Science should keep out of political quarrel," he says. "We do need to work together, and I know that our Russian colleagues feel much the same way."

But worsening international relations could affect ongoing global warming negotiations,

POOL/REUTERS

**RNAi**

Experimental drug may cause blindness, scientists worry

www.nature.com/news

WWW.FREEMAGES.CO.UK



says David Victor, a law professor and energy expert at Stanford University in Palo Alto. "If the Russian relations with the West sour, that presumably will make Russia less willing to sign on to international agreements that they don't see in their national interest," he says. More broadly, he thinks that Russia might also reconsider Western investment in its old industrial and new high-tech sectors.

Meanwhile, Russian scientists involved in collaborations with the West hope that the current tensions will not lead to restrictions in terms of mobility and scientific exchange. "If that happens," says Konstantin Severinov, a molecular biologist at the Institute of Gene Biology in Moscow, "there will be a long-lasting negative effect on Russian science, as it already has strong isolationist tendencies."

Russian researchers have lately experienced a resurgence of state control over their work. For example, Severinov says that he was recently approached by a 'curator' with the Russian Federal Security Service who inquired about his US citizenship, about what exactly he was doing in Russia and in the United States in terms of science, and whether he had a security clearance. "This could be a sign of the times," he says.

If the United States and other countries were to further restrict their visa policies, the isolation of Russian scientists will only increase, says Mikhail Feigel'man, a physicist at the Landau Institute of Theoretical Physics in Moscow. "That would be the most stupid step the West could take," he says.

Quirin Schiermeier, with additional reporting by Jeff Tollefson and Eric Hand

Physicists aflutter about data photographed at conference

An Italian-led research group's closely held data have been outed by paparazzi physicists, who photographed conference slides and then used the data in their own publications.

For weeks, the physics community has been buzzing with the latest results on 'dark matter' from a European satellite mission known as PAMELA (Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics). Team members have talked about their latest results at several recent conferences (see *Nature* 454, 808; 2008), but beyond a quick flash of a slide, the collaboration has not shared the data. Many high-profile journals, including *Nature*, have strict rules about authors publicizing data before publication.

It now seems that some physicists have taken matters into their own hands. At least two papers recently appeared on the preprint server arXiv.org showing representations of PAMELA's latest findings (M. Cirelli *et al.* <http://arxiv.org/abs/0808.3867>; 2008, and L. Bergstrom *et al.* <http://arxiv.org/abs/0808.3725>; 2008). Both have recreated data from photos taken of a PAMELA presentation on 20 August at the Identification of Dark Matter conference in Stockholm, Sweden.

"We had our digital cameras ready," says Marco Cirelli, a theorist at the Institute of Theoretical Physics in Gif-sur-Yvette, France, and one of those who took pictures. The preprints fully acknowledge

the source of the data and reference the presentation photographed.

PAMELA has been attracting such interest because it has reportedly seen an excess of high-energy positrons in space. Those positrons could stem from the collision and annihilation of dark-matter particles, which could make up most of the mass of the Universe. If the data hold up, they would be the most direct clue yet to the nature of dark matter.

The satellite's finding comes at a time when theoretical physicists are desperate for dark-matter data to test their ideas against. "There hasn't been much progress," says Adam Falkowski, a theorist at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. "The hunger for new results in the community is big."

Piergiorgio Picozza, PAMELA's principal investigator and a physicist at the University of Rome Tor Vergata, says he is "very, very upset" by the data being incorporated into a publication. But Cirelli maintains that he and others have done nothing wrong. "We asked the PAMELA people [there], and they said it was not a problem," he says.

Photography or videotaping of conference presentations is common in some fields, such as biology, but is relatively rare in physics. Falkowski says he can't recall another case. Still, he says, "I personally don't find anything wrong with it."

Geoff Brumfiel



F. CHMURA/ALAMY

**RNAi**

Experimental drug may cause blindness, scientists worry

www.nature.com/news

WWW.FREEMAGES.CO.UK



says David Victor, a law professor and energy expert at Stanford University in Palo Alto. "If the Russian relations with the West sour, that presumably will make Russia less willing to sign on to international agreements that they don't see in their national interest," he says. More broadly, he thinks that Russia might also reconsider Western investment in its old industrial and new high-tech sectors.

Meanwhile, Russian scientists involved in collaborations with the West hope that the current tensions will not lead to restrictions in terms of mobility and scientific exchange. "If that happens," says Konstantin Severinov, a molecular biologist at the Institute of Gene Biology in Moscow, "there will be a long-lasting negative effect on Russian science, as it already has strong isolationist tendencies."

Russian researchers have lately experienced a resurgence of state control over their work. For example, Severinov says that he was recently approached by a 'curator' with the Russian Federal Security Service who inquired about his US citizenship, about what exactly he was doing in Russia and in the United States in terms of science, and whether he had a security clearance. "This could be a sign of the times," he says.

If the United States and other countries were to further restrict their visa policies, the isolation of Russian scientists will only increase, says Mikhail Feigel'man, a physicist at the Landau Institute of Theoretical Physics in Moscow. "That would be the most stupid step the West could take," he says.

Quirin Schiermeier, with additional reporting by Jeff Tollefson and Eric Hand

Physicists aflutter about data photographed at conference

An Italian-led research group's closely held data have been outed by paparazzi physicists, who photographed conference slides and then used the data in their own publications.

For weeks, the physics community has been buzzing with the latest results on 'dark matter' from a European satellite mission known as PAMELA (Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics). Team members have talked about their latest results at several recent conferences (see *Nature* 454, 808; 2008), but beyond a quick flash of a slide, the collaboration has not shared the data. Many high-profile journals, including *Nature*, have strict rules about authors publicizing data before publication.

It now seems that some physicists have taken matters into their own hands. At least two papers recently appeared on the preprint server arXiv.org showing representations of PAMELA's latest findings (M. Cirelli *et al.* <http://arxiv.org/abs/0808.3867>; 2008, and L. Bergstrom *et al.* <http://arxiv.org/abs/0808.3725>; 2008). Both have recreated data from photos taken of a PAMELA presentation on 20 August at the Identification of Dark Matter conference in Stockholm, Sweden.

"We had our digital cameras ready," says Marco Cirelli, a theorist at the Institute of Theoretical Physics in Gif-sur-Yvette, France, and one of those who took pictures. The preprints fully acknowledge

the source of the data and reference the presentation photographed.

PAMELA has been attracting such interest because it has reportedly seen an excess of high-energy positrons in space. Those positrons could stem from the collision and annihilation of dark-matter particles, which could make up most of the mass of the Universe. If the data hold up, they would be the most direct clue yet to the nature of dark matter.

The satellite's finding comes at a time when theoretical physicists are desperate for dark-matter data to test their ideas against. "There hasn't been much progress," says Adam Falkowski, a theorist at CERN, Europe's particle-physics laboratory near Geneva, Switzerland. "The hunger for new results in the community is big."

Piergiorgio Picozza, PAMELA's principal investigator and a physicist at the University of Rome Tor Vergata, says he is "very, very upset" by the data being incorporated into a publication. But Cirelli maintains that he and others have done nothing wrong. "We asked the PAMELA people [there], and they said it was not a problem," he says.

Photography or videotaping of conference presentations is common in some fields, such as biology, but is relatively rare in physics. Falkowski says he can't recall another case. Still, he says, "I personally don't find anything wrong with it."

Geoff Brumfiel



F. CHMURA/ALAMY

SPECIAL REPORT

The next Google



Ten years ago this month, Google's first employee turned up at the garage where the search engine was originally housed. What technology at a similar early stage today will have changed our world as much by 2018? *Nature* asked some researchers and business people to speculate — or lay out their wares. Their responses are wide ranging, but one common theme emerges: the integration of the worlds of matter and information, whether it be by the blurring of boundaries between online and real environments, touchy-feely feedback from a phone or chromosomes tucked away on databases.

Bill Buxton

Principal researcher, Microsoft,
Toronto, Canada

ELECTRONIC PAPER

I subscribe to Melvin Kranzberg's second law of technology: invention is the mother of necessity. Although technologies are created to fulfil needs, each also creates them; the next generation of technologies will deliver the promises of what we already have.

The history of communication technologies over the past century tells me that anything that's going to impact on the next ten years is going to be ten years old already. (The components that made Google possible ten years ago were already there ten years earlier, with the creation of the

web.) One prime candidate is electronic paper, displays that are as easy to view in ambient light conditions as paper and that consume hardly any power. It started with E Ink a decade ago; now we are seeing it in devices such as Amazon's Kindle, which I would say has not yet

matured but has certainly reached late adolescence. Kindle and other readers are really like the Ford Model T in terms of what will be available in five years.

I think with this technology will come a dramatic change in our attitude towards paper. Our attachment to paper and books is wonderful, charming and quite understandable. I can't stand reading stuff on my computer. But this technology will make us question whether we can really afford the 500,000 trees that are consumed by publishing and newsprint in North America each week.

Vincent Hayward

Professor of engineering, Pierre and Marie Curie University, Paris, France

HAPTICS

Ten years ago, if you mentioned the word 'haptics' most people would think you were talking about some form of liver disease. Interfaces that provide tactile feedback have been in an innovator-driven 'push' mode; they have been technologically challenging, expensive and restricted to niches. Now there is a public pull, thanks to the spread of touch-screen devices. The objective is to make the interface more intuitive and less reliant on vision — something you can use without looking at it. Haptics makes that possible.

Two or three mobile-phone manufacturers have products on the market with haptic features, and some car companies are doing the same. The feedback acts like an acknowledgement, so you can feel when an onscreen button has been pressed. But also there is something more basic. As animals we operate on the basis of anticipation. Visual interfaces reduce our ability to anticipate because we are touching something that is not there; there is no anticipated sensation and the sensory consequences to our movements are unsatisfactory. Haptic feedback gives us what our minds anticipate; it completes the control loop.

Right now haptic displays are mostly

capable of creating only single isolated sensations of contact, or of toggling through menus. But texture, shape and 'compliance' will become more refined and affordable. A dry, flat screen will be able to simulate the feel of fur or wetness.



ILLUSTRATIONS: N. SPENCER

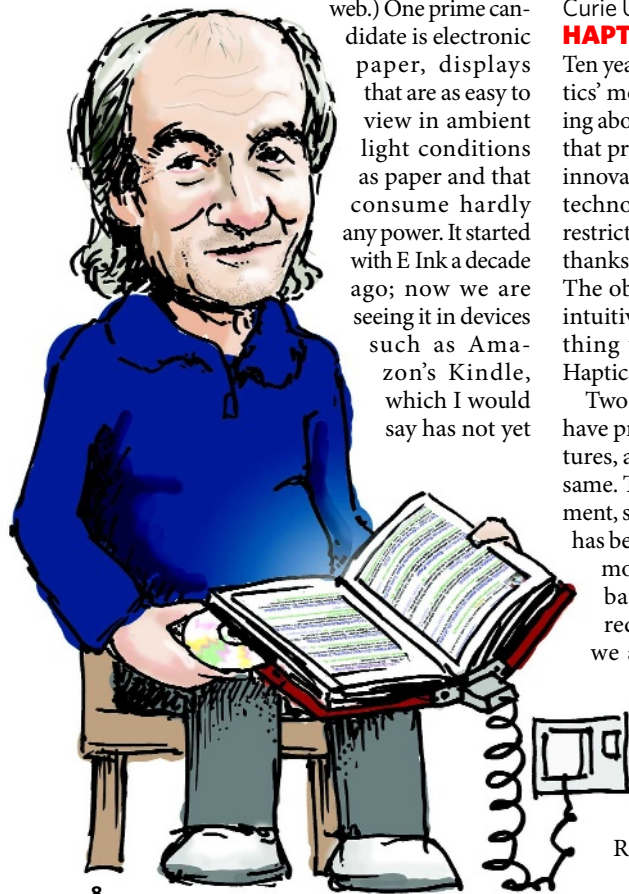
Ian Pearson

Futurizon consultancy, Ipswich, UK

VIDEO VISORS

We're crying out for technology that will allow us to combine what we can do on the Internet with what we do in the physical world. That's why the Nintendo Wii has been so successful. One technology that springs to mind is the video visor, which gives you a computer image superimposed over the world around you.

These have been around for a few years, but they currently have pretty low resolution. The resolution will improve and the cost will come down; at the same time demand will grow because the visors can provide information to people on the move. People have their iPhones and Blackberries with lots of data and functions but they want bigger displays. Wearing visors may seem odd at first, but then people used to stand out when mobile phones and Bluetooth





headsets first came out. Now everyone uses them.

When you start to combine visor graphics with more accurate global-positioning data, as will be provided by the European Galileo satellites, you can overlay online information onto the world around you. So as you're walking down a busy city street you will be able to see reviews of shops and restaurants, adverts for services, other people who have similar interests to you, or whatever.

When you are wearing a visor your surroundings can have a completely different appearance: a burger restaurant can look like a giant burger without flouting planning laws. You could be seen as your Second Life virtual avatar. Or Johnny Depp, or Claudia Schiffer. You get the best of both worlds.

Leo Kärkkäinen

Chief visionary, Nokia Research Center, Espoo, Finland

PRODUCTS WITH MEMORIES

Ordinary products are going to have memories that store their entire history from cradle to grave, and that consumers can easily access.

Radio-frequency identification tags are a good option because they are already widely used to track inventory and to control theft. They are cheap and can be powered by an outside power source, such as the radio signal from the device being used to read them. But there may be another enabling technology that wins out.

Near-field communication systems already allow a phone to be used like a smart card for a travel pass or as an electronic wallet to pay for goods. If that technology can talk to the things you buy, as well as the systems through which you pay for them, it will enable consumers to choose not to buy goods that are unhealthy, allergenic, have used environmentally unfriendly methods or employed child labour.

As with many technologies, it could potentially be used for bad purposes; we have to ensure that privacy functions are built in to the system to put the consumer in control of whether they want to be tracked.

Helen Greiner

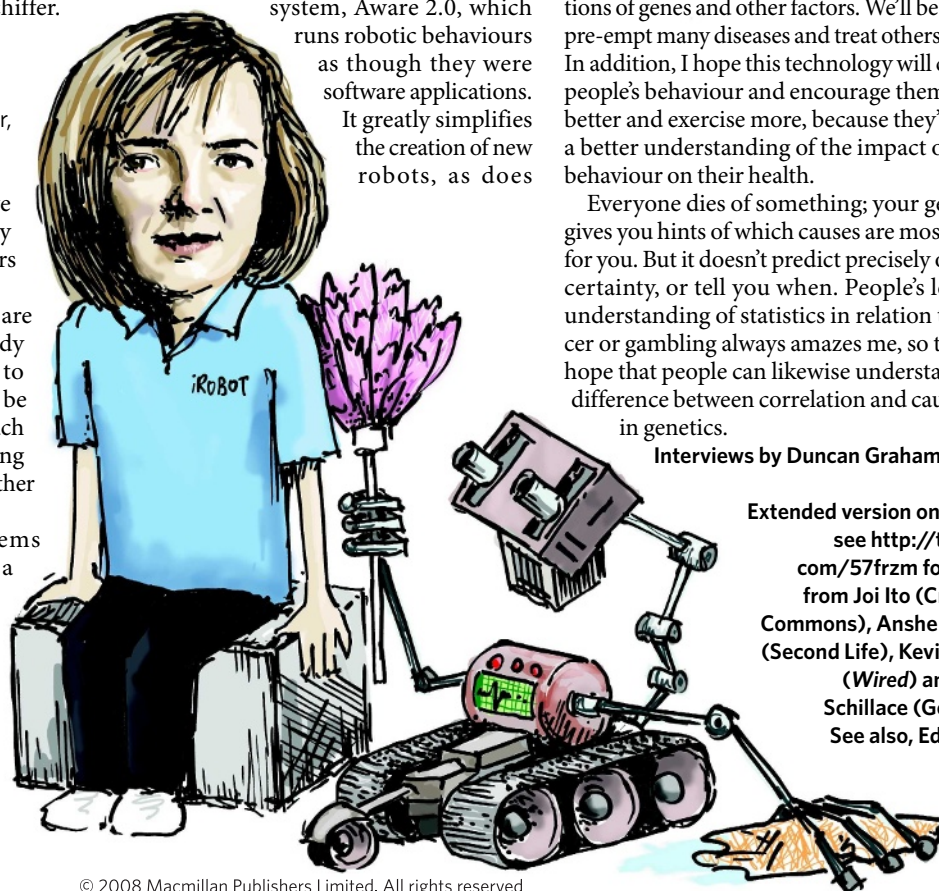
Chairman and co-founder, iRobot, Burlington, Massachusetts

AUTONOMOUS ROBOTS

Others have said it before, but I now think it's a safe bet to say that within the next ten years robots will become a lot more commonplace. The key is autonomy. Unless a robot has 'mission based' autonomy, it needs to be controlled by a human; this makes sense for something critical such as a military operation, but is often just a waste of time. Now we're seeing robotic agents that can go out and act on their own: ploughing fields, mowing lawns or cleaning offices. Increasingly autonomous robots will be capable of more complex and sophisticated behaviours, taking on more complex chores and tasks in agriculture, construction, logistics, care of the elderly, the military and the home.

To get autonomy you need perception of the environment, an intelligent software architecture, a physical system or body and behaviours. Our Roomba vacuum cleaner is an example of autonomy with all these features.

We've now created a sort of robotic operating system, Aware 2.0, which runs robotic behaviours as though they were software applications. It greatly simplifies the creation of new robots, as does



HAVE YOUR SAY

Comment on this or any of our news stories, online.
www.nature.com/news

modularity in the mechanical design, the perceptive systems and the components of intelligence. That makes it possible to build on past successes; once you have developed a navigation behaviour, for example, it can be used in other platforms.

Esther Dyson

Investor in for-profit and not-for-profit start-ups, New York

GENETIC INFORMATION

I'm on the board of 23andMe of Mountain View, California, which makes genetic information accessible to its owners — and lets them share it for research if they want to.

For now, 23andMe looks only at common genetic variations, which mostly show risk factors — there are only a few conditions for which a genetic anomaly indicates almost 100% risk, and even then you might not know the timing or intensity. Our service, which costs US\$1,000, will become cheaper as the cost of the information processing, the chemistry and the imaging technology comes down and can be spread over a broader base of customers.

The first users are mostly benefactors; later users will be beneficiaries. As hundreds of thousands, and eventually millions, of people take part, the genetic information collected will enable us to know so much more through data mining, combined with analysis of the interactions of genes and other factors. We'll be able to pre-empt many diseases and treat others better. In addition, I hope this technology will change people's behaviour and encourage them to eat better and exercise more, because they'll have a better understanding of the impact of their behaviour on their health.

Everyone dies of something; your genome gives you hints of which causes are most likely for you. But it doesn't predict precisely or with certainty, or tell you when. People's level of understanding of statistics in relation to soccer or gambling always amazes me, so there is hope that people can likewise understand the difference between correlation and causation in genetics.

Interviews by Duncan Graham-Rowe

Extended version on line — see <http://tinyurl.com/57frzm> for picks from Joi Ito (Creative Commons), Anshe Chung (Second Life), Kevin Kelly (Wired) and Sam Schillace (Google). See also, Editorial, page 1.

The next Google

Additional text for article doi:10.1038/455008a

Joi Ito

Co-founder of Infoseek Japan and chief executive of Creative Commons, Tokyo, Japan

OPEN CONTENT MANAGEMENT

The next big thing will come from connecting people and ideas together with a Google-like simplicity — making Wikipedia, Facebook and all sorts of other things completely seamless. It sounds obvious and yet it's hard to imagine. But then, before Google it was hard to imagine what search could be like. Before Tim Berners-Lee it was hard to imagine the web.

I think that a key part to it will be software that automatically gives attribution for the various parts of content we access and share. People want to share content with each other, but the infrastructure and legal framework makes it more difficult than it should be. Legal friction is holding back a lot of creativity. If you have software that works out who owns what for you and gives credit where it is due, and if it can support all different kinds of content, then you start to have a network that enables a great deal more creativity.

Anshe Chung

Avatar of Ailin Graef, the first person to achieve a net worth of more than a million dollars from profits earned in a virtual world, Second Life

THREE-DIMENSIONAL ENVIRONMENTS

I think that the physical and virtual will merge more and more over the next decade as three-dimensional (3D) environments become increasingly easy to use through normal browsers and mobile phones.

These 3D scenes will represent real people and real places — things of value. When I enter a 3D scene and know it is an up-to-date copy of Manhattan, and interact with other

users who are either virtually present or even physically located in the real place, it becomes far more meaningful than a fantasy world or a game.

Social worlds such as Second Life have managed to create 3D communities of hundreds of thousands of people, but accepting the simple avatars and the environment requires learning and effort. There are several technologies that could help realize this. The first is the computer graphics with which to create photo-realistic images. The second is the means to capture huge parts of the physical world and add them to the 3D world. Companies such as Google and Microsoft have already started doing this using satellite images and huge amounts of imagery in cities, with users contributing by adding data and metadata. The third technology is representations of people that bring them into the space mentally and allow them to interact with it better.

Kevin Kelly

Founding executive editor, *Wired* magazine, Pacifica, California

THE SEMANTIC WEB

The semantic web is very difficult to explain because there's nothing really to look at. Google had a sparse homepage — the semantic web doesn't have anything at all. But I think the total effect of it will be at least equal to that of Google.

The idea is that if everything on the web was described and reduced down to a noun, verb and predicate, as in a language, computers could 'read' the web. It would have meaning. Then machines can do a lot of the things normally done by people because they can suddenly read information. If you want to book a taxi to the airport, the semantic web gives a machine the ability to know certain things: it will know your flight times, that there are road

works on the way to the airport, which cab firm you prefer, and so on. A second-order effect would be that the information would come to you, rather than you go to it.

But getting there is a chicken-and-egg problem. Hand coding is very laborious: the initial benefits are small. So until there's a critical mass it's difficult to persuade people to do it. The breakthrough for search engines was page ranking. With the semantic web the tipping point could come from something like an automated parser, which codes the meaning of content automatically. There are some websites, such as Twine, that are beginning to do it.

Sam Schillace

Google, Mountain View, California

BETTER BROWSERS

Prior to Google, everyone said search was done. But the point was that search could have been a lot better. The same is true of browsers today.

On the web, simplicity matters more than completeness — the platform needs to be simple, ubiquitous and good enough. The browser is that platform. It means any screen you look at can be a window into your own personal, private cloud of information. I use three different computers every day but don't worry which of them a particular file, picture or e-mail is on, because they are online and my browser can find them.

The current generation of browsers can already run some pretty sophisticated applications without having to install software, and it's starting to extend to mobile devices too. The next generation of browsers, and the web applications that run on them, will make communication and collaboration even more transparent and let me focus on what I really want to do — connect with the person at the other end and get work done together. It will

**RADIOISOTOPES**

Reactor shut-downs delay cancer treatments

www.nature.com/news

PUNCHSTOCK

Mathematical biology centre launched

A new US institute that will be part funded by the Department of Homeland Security (DHS) will aim to become the world centre for collaboration between mathematicians and biologists.

The National Institute for Mathematical and Biological Synthesis (NIMBioS) will bring mathematical approaches to problems across biology, with a particular focus on modelling the dynamics of animal disease. The National Science Foundation (NSF) will, within a week, announce that the University of Tennessee in Knoxville has beaten 18 other proposals to host the centre.

"We want to become the place people think of first in linking mathematics and biology," says the institute's director Louis Gross, a mathematical ecologist at the University of Tennessee. "Mathematical biology has traditionally been one little corner of biology. We want to move it to a central role."

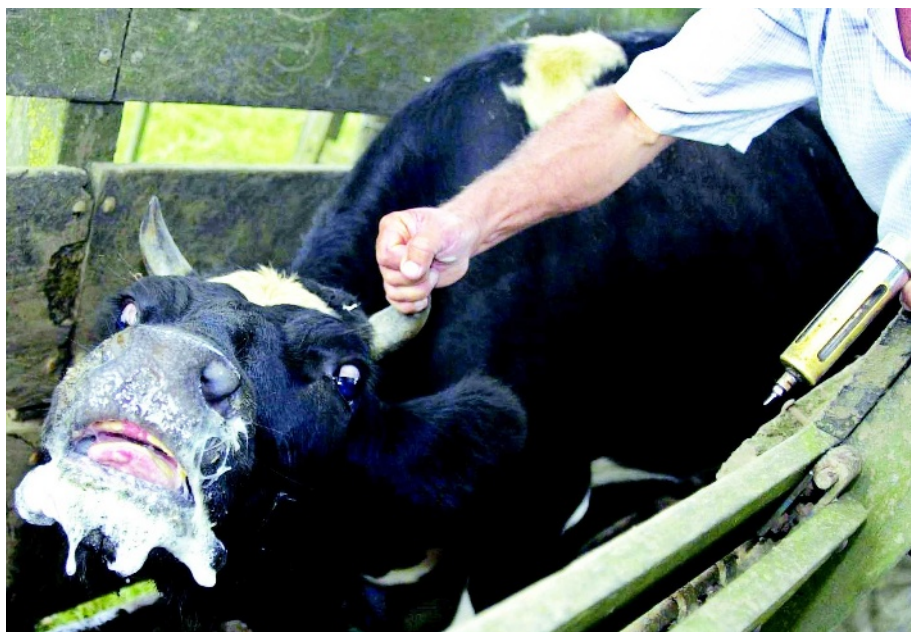
The institute's creation reflects the growing strength of mathematical biology, and growing concern about the potential impact of animal diseases on agriculture and human health, as shown by outbreaks worldwide of foot-and-mouth disease, avian influenza and SARS. Four-fifths of emerging human diseases cross over from animal infections, says Tam Garland, branch chief for agricultural security at the DHS.

"A whole series of events has raised concerns within the federal government that this is something we need to be aware of," says Samuel Scheiner, programme director in the NSF's division of environmental biology. Modelling has already proven its ability to predict and help control disease outbreaks, he says — for example, understanding the population cycles of hantavirus, which crosses from small mammals to humans, has allowed disease peaks in the southwest United States to be forecast and reduced.

"Modelling is a decision tool," says Garland. "What we're supporting with this centre is fundamental research and growing the next generation of researchers." A long-term goal, she adds, is to be able to distinguish natural outbreaks from possible deliberate release.

Besides ecology and evolution, which already have strong mathematical components, NIMBioS aims to bring mathematics to parts of biology that it has so far had little impact on, such as development and immunology.

The US government has pledged NIMBioS



M. MENDEZ/AFP/GETTY IMAGES

The new institute will focus on modelling animal diseases such as foot and mouth.

US\$16 million over five years. Of this, \$11 million will come from the NSF and \$5 million from the DHS. The aim is to do basic research, not provide, say, rapid-response advice on vaccination or culling in response to a disease outbreak.

NIMBioS plans to supplement its budget partly through contract work, providing simulations and analyses for land managers. The institute has also signed up IBM and ESRI, a geographical software company based in Redlands, California, as industry partners, although neither has contributed funding so far. Gross anticipates an annual operating budget of about \$5 million.

Most of the core funding will be spent on about a dozen postdoctoral positions, and on working groups bringing together 8–15 researchers to study a particular problem in a series of two or three approximately week-long meetings spread over a couple of years.

The eight or nine groups planned in the first year include investigations of the links between the mathematics of invasive species and cancer metastasis; the dynamics of social networks in animals; and modelling the spread of pseudorabies among feral pigs in the southern United States.

This approach mimics that of other NSF-funded centres such as the National Center for

Ecological Analysis and Synthesis (NCEAS), founded in 1995 and based at the University of California, Santa Barbara.

"Other institutes like this have had a tremendous global impact," says Alan Hastings, a theoretical ecologist at the University of California, Davis. Work at the NCEAS has been important in giving applied ecology a scientific underpinning, such as in the design of marine reserves, he says. "The NCEAS changed the way people do ecology."

A 20-strong international governing board will review NIMBioS proposals for quality, and to avoid duplicating the work of existing centres such as the NCEAS and the Mathematical Biosciences Institute at Ohio State University in Columbus.

Mathematical biology is growing worldwide, but European groups tend to be "dispersed and specialized," says Wolfgang Alt, a theoretical biologist at the University of Bonn, Germany, and president of the European Society for Mathematical and Theoretical Biology.

Japan is in a similar position to Europe, says Nanako Shigesada, a theoretical ecologist at Doshisha University in Kyoto and president of the Japanese Society for Mathematical Biology. "Having a research institute covering all areas of mathematical biology, including ecology, evolution, developmental and cellular and subcellular processes is very important," she says. ■

John Whitfield

"Other institutes like this have had a tremendous global impact."

THE ENERGY ELECTION



In the first of a special series of election podcasts starting this week, *Nature* gathered an expert panel to discuss how energy and climate issues will play out in the US presidential election. Excerpts:

"The world has made transitions from one type of energy source to another ... in the 75- to 125-year kind of timeframe. We don't have that luxury here. We have to hurry history."

Steve Cochran, Environmental Defense Fund, Washington DC

"However you want to cut it — if we're going to get serious about climate-change policy, we're going to have to change the prices of fossil fuels."

Joseph Aldy, fellow, Resources for the Future and co-director of the Harvard Project on International Climate Agreements, Washington DC

"We need ... to identify in very concrete terms, not just in a sort of warm and fuzzy way, what new investments in the energy sector mean: where those dollars would go, where those jobs would be created, where an auto worker who is currently making an SUV will now be making a hybrid transmission."

Steve Cochran

"If we put too much money into energy R&D over too short a period of time, there is going to be waste."

Richard Newell, professor of energy and environmental economics, Duke University, Durham, North Carolina

"We need to have a diversified portfolio of R&D [and] we shouldn't pick just one winner. Having said that, let me pick a winner right now ... carbon capture and storage."

Joseph Aldy

To hear the full discussion, chaired by our columnist David Goldston (see page 15), visit www.nature.com/nature/podcast. Future podcasts in this series will cover biomedical research and innovation policy.



Republicans at odds over human embryo research

By changing one little word, the committee drafting the Republican 2008 election platform calls for banning all human embryo research in the United States, whether publicly or privately funded.

John McCain, the presumptive Republican presidential nominee, is under no obligation to follow the party platform — which is a statement of principle with no binding power — but the change highlights the already noticeable contrast between him and the official party position. Although his running mate, Governor Sarah Palin of Alaska, opposes human embryonic stem-cell research, McCain has twice voted to loosen restrictions on federal funding of the work.

On 27 August, the Republican Platform Committee approved an amendment by Mary Summa of North Carolina, one of its 100 or so delegates. It changed "and" to "or" so that the platform now calls for a ban on "the creation of *or* experimentation on human embryos for research purposes" (emphasis added). The change won final approval during the Republican convention this week in St Paul, Minnesota.

By unlinking the creation of embryos from experimentation on them, the amendment effectively proposes banning a huge swathe of research — from attempts to improve preservation of frozen embryos at *in vitro* fertilization clinics to the privately financed creation of new stem-cell lines.

The Republican National Committee last week declined to comment on the platform, saying it was not yet official.

The change highlights a rift between social conservatives and Republican moderates such as Michael Castle of Delaware, a Republican member of the House of Representatives and a leading supporter of lifting the funding restrictions. The change, Castle says, "was drafted by people who don't even understand the advances that have been made in embryonic stem-cell research and its future potential".

But David Christensen, the leading lobbyist on embryo-related issues at the Family Research

Council, a conservative Christian advocacy group in Washington DC, praised the change as "very consistent with the traditional Republican platform that calls for the protection of the dignity of all human life regardless of stage of development".

Even if McCain were to adopt party tenets, the stem-cell restrictions would stand virtually no chance of being enacted by a Democratic-led Congress. Still, says Sean Tipton, director of public affairs at the American Society for Reproductive Medicine in Washington DC, "it adds to a chilling effect on the research. Even a whiff of a prohibition of private work just further curtails researchers', investors' and philanthropists' interests."

"I find it almost inconceivable that they would take such a backwards step at

this point in time," adds Peter Mathers, who chairs the stem-cell subcommittee of the science-policy committee for the Federation of American Societies for Experimental Biology. "Remaining neutral is one thing. Going backwards seems to be very disconcerting."

Barack Obama, the Democratic nominee for president, has, like McCain, voted to lift federal funding restrictions on stem-cell research. Last week, the science advocacy group Science Debate 2008 released answers from Obama on several science-related topics, including a statement on stem-cell research that he favours "responsible oversight of it, in accord with recent reports from the National Research Council".

As with stem-cell research, McCain and his party also diverge on climate change. The proposed platform cautions against "doomsday climate-change scenarios peddled by the aficionados of centralized command-and-control government". Echoing the 2004 platform, it advocates "technology-driven, market-based solutions" to increased atmospheric carbon.

McCain, in contrast, has promised to enact mandatory limits on greenhouse-gas emissions through a cap-and-trade system.

Meredith Wadman



John McCain and running mate Sarah Palin.

J. RAEDLE/GETTY IMAGES

DNA databases shut after identities compromised

Several DNA databases run by the US National Institutes of Health (NIH) in Bethesda, Maryland, the Wellcome Trust in London and the Broad Institute in Cambridge, Massachusetts, were closed to public access last week after researchers showed it is possible to extract the supposedly confidential identities of the patients involved. The databases list the frequencies of small DNA variations called single nucleotide polymorphisms (SNPs) from patient groups.

In the August issue of *PLoS Genetics*, Nils Homer and his colleagues describe a method to mine individual SNP profiles from complex mixtures, even if the person's DNA is only 0.1% of the total. The method could be useful for ensuring patients are not listed twice when scientists combine data sets, as well as in forensic science.

The NIH has not identified any patient privacy violations, and points out that to identify a particular patient, one would need his or her genetic profile. Researchers will now have to apply for data access at the individual level, as they do for study data.

Bubble-fusion researcher loses misconduct appeal

Nuclear engineer Rusi Taleyarkhan has been stripped of his named professorship at Purdue University in West Lafayette, Indiana, following the results of a misconduct inquiry into his bubble-fusion research.

Having lost an appeal against the university's misconduct ruling, Taleyarkhan is banned from having graduate students for three years, and loses the title 'Arden L. Bement Jr Professor of Nuclear

Engineering', along with annual resources of \$25,000 that come with it. He will, however, remain on the faculty and have his situation reviewed in three years.

According to an investigation report released on 18 July, Taleyarkhan's misconduct involved two falsifications of the research record. In a recent e-mail to *Nature*, Taleyarkhan denies both charges, calling the findings "grossly inappropriate" and adding, "the two allegations for which misconduct was concluded have nothing to do with the science of bubble nuclear fusion". He also questions the sanctions, given that "a duly constituted committee in 2006 looking at these same two issues" exonerated him.

For a longer version of this story, see <http://tinyurl.com/5mqyhn>

'No pollution effects' from Chinese chemical explosion

An explosion that killed 20 people last week at a chemical plant in Yizhou, in the southern Chinese province of Guangxi, poses no further threat, according to preliminary surveys by the Chinese health ministry.

The cause of the 26 August explosion is under investigation. Liu Xiongmin, an engineer specializing in disaster management at Guangxi University in Nanning, says the explosion may have been caused by a chemical leak coupled with very hot temperatures that day, which reached 36 °C. The plant produced polyvinyl acetate, carbide and vinyl acetate.

The explosion destroyed the five-storey plant as well as nearby houses. Liu says that of the bulk chemicals used at the plant, ethanol was mostly burned in the explosion and methyl alcohol does not seem to have escaped in large amounts. Reportedly, none of the 60 people injured in the blast had toxic reactions.



Alaska's polar bears trigger lawsuit from industry

Five industry groups are disputing the US decision to list the polar bear as a threatened species. Last week, they filed suit in federal court to attempt to change the text of the listing that says business projects in Alaska — but no other state — must undergo reviews of their greenhouse-gas emissions.

The American Petroleum Institute, the US Chamber of Commerce, the National Mining Association, the National Association of Manufacturers and the American Iron and Steel Institute say the listing unfairly singles out Alaskan businesses.

It is the latest legal challenge to the May listing of the polar bear. The state of Alaska, led by governor and presumptive vice-presidential candidate Sarah Palin, has sued the government over the decision, saying it harms oil and gas exploration in the state. For its part, the environmental group the Center for Biological Diversity, which pushed for the original listing, is suing to get the polar bear upgraded from 'threatened' to 'endangered'.

'YouTube for test tubes' to be listed on PubMed

The *Journal of Visualized Experiments* (*JoVE*) has announced that its online video protocols will be indexed in the popular US National Library of Medicine repositories MEDLINE and PubMed.

Founder and chief executive Moshe Pritsker views the MEDLINE-PubMed listing as a sign that the scientific community has accepted video-based publications. "It was a very important decision for us, and for scientific publishing," he says.

Since *JoVE* was founded in 2006 with support from an angel investor, the journal has published more than 200 videos, most produced by professional videographers. It aims to improve the reproducibility of scientific results by using videos to clarify subtle experimental details. The journal was itself an experiment in video publishing and remains the only video-based scientific journal.

Mars rover climbs out of crater to focus on plains

Almost a year after it drove into the Victoria Crater on Mars, NASA's Opportunity rover last week made its way back out. Opportunity left by the same route it entered by, allowing it to study its old tracks (pictured) for any signs of changes during the past 11 months.

During its time in the 800-metre-wide, 70-metre-deep crater, the rover took detailed photographs of rocky outcrops and conducted chemical analyses that suggest the crater was once soaked in water. Opportunity will next study rocks strewn across the surrounding plains. Scientists hope that these represent several types of rock scattered by earlier impacts.

Opportunity has spent more than 1,600 martian days exploring and is showing its age. Its robotic arm is kept permanently extended owing to fears about the health of its 'shoulder' motor, and engineers recently noted an electric-current spike similar to that seen in Opportunity's twin rover Spirit shortly before it lost the use of one of its wheels.



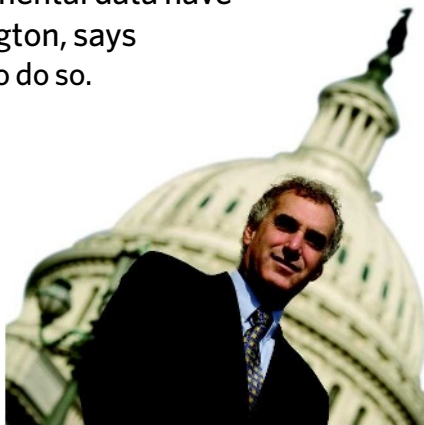
Data wrangling

Collecting and releasing environmental data have stirred up controversy in Washington, says **David Goldston**, and will continue to do so.

Data sound like a grey, non-partisan and unemotional topic for political discussion. But decisions on what data to collect and release for use in research or policy-making are hardly neutral in their impact. This may be clearest in the arena of environmental policy, where hard-fought disputes over the collection and dissemination of data frequently break out. Indeed, perhaps the only thing politicians agree on about environmental data is that more data are always better — in theory.

Maybe only in theory. One of the continuing debates has been over how much data collection to fund. Although advocates on both right and left frequently call for more data — if sometimes just to delay decision-making — sensors that measure such matters as air and water quality are often the first items to be cut when budgets get tight. And efforts to develop a set of environmental indicators that would be regularly updated — something akin to economic statistics — have not got very far. This can be seen in the second *State of the Nation's Ecosystems* report (www.heinzcenter.org/ecosystems), released in June by the Heinz Center, a private think tank in Washington DC. Although the report includes many environmental measurements, it is chock-full of lists of subject and geographical areas for which few if any data exist. Indeed, a companion volume urges a significant expansion of federal funding for environmental indicators.

Such an expansion, though warranted, seems unlikely any time soon. Politicians like to talk about the need for more data, but it is rarely anyone's top priority. In fact, when the Bush administration responded to the Heinz report by announcing that it would put more money into indicators — a move only symbolic this late in its term — one environmental organization took the White House to task for spending money on measuring pollution rather than cleaning it up. Broad data collection not connected to any single controversy isn't very sexy and must compete with many related activities presumed to have more immediate impact (although it may be hard to tell without the data). Even when instrumentation is regularly funded, as some kinds of satellites are, money is often lacking to maintain the data or to make them sufficiently accessible or digestible.



PARTY OF ONE

And if data collection and processing were to be institutionalized, another ongoing debate would have to be resolved — how insulated the operations should be from politics. For years, there has been talk of establishing a Bureau of Environmental Statistics (BES), which would not only gather data but also analyse them. Data are never as straightforward a matter as they seem; just deciding what information to collect involves judgements about what's important. You don't measure, say, pesticide levels in food unless you think they're a problem. And deciding that a problem exists is different from deciding what to do about it. The frequently heard claim that 'the data speak for themselves' has to be one of the most misleading sentences in the English language.

Still, a statistical agency needs to be free from political manipulation to have any credibility. Around 2002, I was involved in lengthy, closed negotiations on Capitol Hill between moderate and conservative Republican congressmen interested in setting up a BES. But the effort was eventually scuttled when the Bush administration rejected all proposals to keep the agency at arms length from politics, arguing that the heads of all major agencies should be responsible to the president.

Even without a BES, the US government releases a lot of environmental data. Much of this is information to determine compliance with regulations, but increasingly just making data available is seen as a way to encourage companies to clean up their operations. The model for such efforts is the Toxic Release Inventory (TRI), established by Congress in 1987, which requires companies to publicly report their annual emissions of certain

chemicals. The TRI has resulted in substantial cutbacks in emissions as companies try to 'green' their reputations. Bush administration proposals to save industry money by reducing the frequency of reporting or the number of companies required to report have met with widespread opposition. But expanding reporting has also been controversial. Chemical firms have resisted reporting what chemicals they use (as opposed to release), arguing that doing so would reveal too much about their operations to competitors and would provide too much information for potential terrorists.

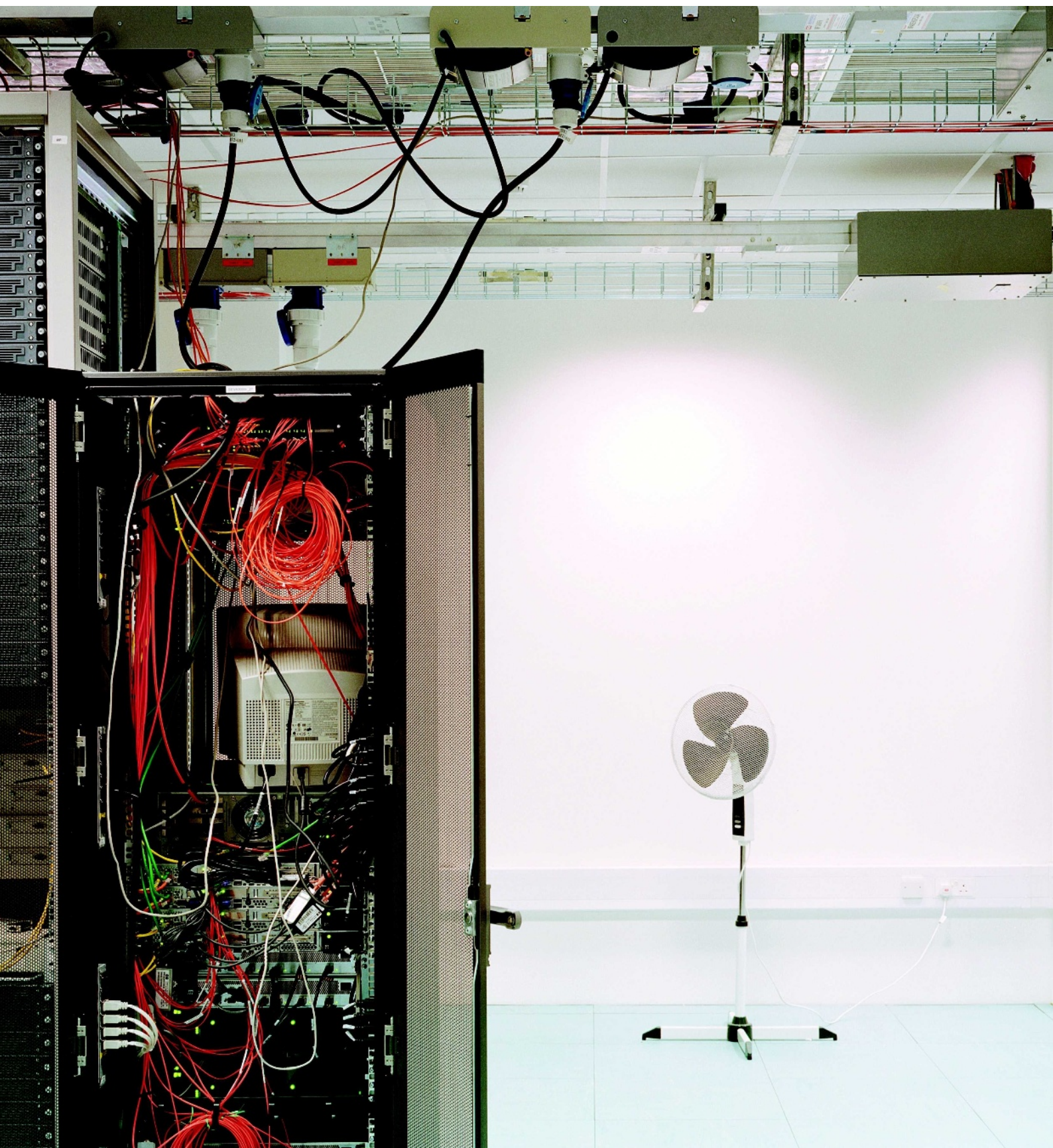
Industry has also been concerned about alleged inaccuracies in data on government websites. The Data Quality Act, enacted in 2000, requires federal agencies to enable private parties to challenge the accuracy of information being disseminated by the government. The law has been anathema to environmental groups, which have seen it as a way to stymie regulation. And it has been primarily invoked by corporations questioning studies that raise alarms about their products. The statute was written after academic researchers declined to release the raw data behind epidemiological studies that were being used to toughen clean-air regulations in 1997, citing privacy concerns.

Data sharing by individual, non-governmental scientists has increasingly become a topic for public debate. Charging that a scientist has been unwilling to share data is a good way for politicians to raise suspicions about someone's work, especially when the work itself is too technical to be easily evaluated by laymen. But different fields have different mores about data sharing, and the issue is not clear-cut. For example, Michael Mann, the author of a controversial study on the history of Earth's temperature — the 'hockey stick' graph — was attacked by conservatives for not sharing his data. But what he had actually held close was not his data, but his computer code, which he claimed, with government backing, was his intellectual property. He did eventually release the code.

In the political sphere, talking about the need for public data is always a good way to sound objective and above-the-fray. But data are a complicated matter, and by themselves rarely resolve an underlying controversy or problem. Nonetheless, the siren song of data has a long history. When the idea of publicly collecting and releasing information was in its infancy, an eighteenth-century Enlightenment thinker proclaimed that statistics and tyranny were incompatible. That turned out to be untrue, too. ■

David Goldston is the former chief of staff of the House Committee on Science. Reach him at partyofone@gmail.com.

See Editorial, page 1.



WELCOME TO THE PETACENTRE

What does it take to store bytes by the tens of thousands of trillions? **Cory Doctorow** meets the people and machines for which it's all in a day's work.

Ten seconds after I stepped into the roar of the data centre at the UK Wellcome Trust Sanger Institute, in rural Cambridgeshire, my video camera croaked: CARD FULL. Impossible. That morning, I'd tossed a handful of thumbnail-sized 32-GB memory cards into my pocket, each one good for a couple of hours' worth of high-definition video. Yet this one had filled up in seconds.

I fumbled with my camera while Phil Butcher, the Sanger Institute's head of information technology (IT), politely waited, grinning in the shower of cold air washing down from the air conditioning. It took only a couple of embarrassing seconds to troubleshoot: I'd somehow mixed an old 32-megabyte card in with the 32-gigabyte cards. The 32-MB card is only a couple of years old; when I bought it, it probably cost more than the 32-GB cards do today. But it holds one one-thousandth of the data.

That, in coincidental microcosm, is the story I'm here for: the relentless march from kilo to mega to giga to tera to peta to exa to zetta to yotta. The mad, inconceivable growth of computer performance and data storage is changing science, knowledge, surveillance, freedom, literacy, the arts — everything that can be represented as data, or built on those representations. And in doing so it is putting endless strain on the people and machines that store the exponentially growing wealth of data involved. I've set out to see how the system administrators, or sysadmins, at some of the biggest scientific data centres take that strain — and to get a sense of how it feels to work with some of the biggest, coolest IT toys on the planet.

At this scale, memory has costs. It costs money — 168 million Swiss francs (US\$150 million) for data management at the new Large Hadron Collider (LHC) at CERN, the European particle-physics lab near Geneva. And it also has costs that are more physical. Every



watt that you put into retrieving data and calculating with them comes out in heat, whether it be on a desktop or in a data centre; in the United States, the energy used by computers has more than doubled since 2000. Once you're conducting petacalculations on petabytes, you're into petaheat territory. Two floors of the Sanger data centre are devoted to cooling. The top one houses the current cooling system. The one below sits waiting for the day that the centre needs to double its cooling capacity. Both are sheathed in dramatic blue glass; the scientists call the building the Ice Cube.

Blank slate

The fallow cooling floor is matched in the compute centre below (these people all use 'compute' as an adjective). When Butcher was tasked with building the Sanger's data farm he decided to implement a sort of crop rotation. A quarter of the data centre — 250 square metres — is empty, waiting for the day when the centre needs to upgrade to an entirely new generation of machines. When that day comes, Butcher and his team will set up in that empty space the yet-to-be-specified systems for power, cooling and the rest of it. Once the new centre is up, they'll be able to shift operations from the obsolete old centre in sections, dismantling and rebuilding without a service interruption, leaving a new patch of the floor fallow — in anticipation of doing it all again in a distressingly short space of time.

The first rotation may come soon. Sequencing at the Sanger, and elsewhere, is getting faster at a dizzying pace — a pace made possible by the data storage facilities that are inflating to ever greater sizes. Take the human genome: the fact that there is now a reference genome sitting in digital storage brings a new generation of sequencing hardware into its own. The crib that the reference genome provides makes the task of adding together the tens of millions of short samples those machines produce a tractable one. It is



S. NORFOLK

Left: the data centre at the Wellcome Trust Sanger Institute in Cambridge, UK, under development.



what makes the 1000 Genomes Project, which the Sanger is undertaking in concert with the Beijing Genomics Institute in China and the US National Human Genome Research Institute, possible — and with it the project's extraordinary aim of identifying every gene-variant present in at least 1% of Earth's population.

As data pour off the Sanger's new Solexa sequencers, Butcher — a trim, bantam grey-haired engineer with twinkling eyes and laugh-lines — has to see to it that they have somewhere to go. A two-hour Solexa run produces a gigantic amount of information: 320 TB, according to Tony Cox, head of sequencing informatics, a figure he's mentioned to journalists in the past (a print-out on his office door reads: "Oh shit, that's 320 TB!" — Tony Cox, *The Guardian*, 28 February 2008"). The 1000 Genome Project needs to make use of storage and computing capacity at a (currently) impossible density. Luckily for Butcher, 'impossible' is a time-bound notion — if you don't like the compute reality, just wait around a little while and another will be along shortly. His storage density is doubling every year; the 500-GB hard-drives spinning away in his storage array are being phased out by Seagate of Scotts Valley, California, the company that makes them, in favour of a terabyte model.

Finding a place for the data to go is only the beginning. Butcher also has to make sure they can get back out. The Sanger has a commitment to serving as an open-computing facility for the worldwide research community. So it faces what you could call the Google problem: an unpredictable and capricious world that might decide at any moment to swarm in with demands for shedloads of arbitrary-seeming data. Just as a news scandal can conjure a flashmob of millions of net-users to Google's homepage, all searching for 'tsunami' or 'paris hilton', an exciting discovery in genetics sends the whole bioinformatics community to the Sanger's servers, all demanding the same thing.

You can't go far in this world without some sort of comparison to Google, which is the biggest of the big gorillas. How big, though, is not quite clear — and nor is it clear how it manages its flashmobs and other petaproblems. In the ten years since the company's founding, Google's data-serving systems have gone from a set of commodity PCs connected to a hard-disk array built into an enclosure made from Lego bricks

The XS4ALL building in Amsterdam (top left); the back of stacks at XS4ALL (bottom left) and CERN near Geneva (facing page, top); a PetaBox (facing page, bottom); and Tony Cox's door (inset).

to a global system of data farms unmatched by anyone else. Each of those data centres is designed from the foundation up to operate as a single big computer. Google buys components optimized for the kind of operations that relentlessly hammer its servers. It has software for the job such as Google File System (which distributes three copies of each piece of information in a way that makes it easy to recover when the inevitable failure occurs) and Google MapReduce, a system for automatically and efficiently making large data sets amenable to parallel processing. Google's distinguished engineers present papers at learned conferences explaining in detail how it all works. People such as Butcher pay close attention.

But then there's the closed part: all the specific metrics and data-porn that Google considers of competitive significance. The company no longer says how big the model, or copy, of the web's material spread through its data centres is. It doesn't disclose the dimensions or capacity of those data centres. *Nature* wanted me to visit one for this piece, but a highly placed

Googler told me that no one from the press had ever been admitted to a Google data centre; it would require a decision taken at the board level. Which is too bad. But it's not as if the world is bereft of other computer installations with mind-bending requirements.

And at CERN, the Sanger and XS4ALL in the Netherlands, I found myself welcomed into the roaring sanctums of computing, escorted around by sysadmins eager to show off the hellaciously complex, monstrously powerful machines that they've been able to put together and put to use.

Repository of all knowledge

The primary XS4ALL facility at the World Trade Centre near Schiphol Airport in Amsterdam was actually built to house a mighty array of huge telephone switches in 2000. KPN, the then Dutch national telecom company, fitted it out generously, with several months' worth of diesel in subterranean tanks for its uninterruptible power supply's back-up generators, and two independent cooling systems, with one raised two storeys off the ground to flood-proof it (this is the lowlands after all). But telecom deregulation was not kind to KPN, and the switches never came. So now the facility houses XS4ALL, a once-notorious Dutch Internet service provider that has somehow made it bigtime. Hacktic, a collective of hackers, established XS4ALL in 1993 to help cover the costs of the Internet link they had set up. In 1994, KPN shut down all of XS4ALL's lines after the collective published

"The PetaBoxes have the elegant, explosive compactness of plutonium."

C. DOCTOROW

C. DOCTOROW

an article explaining how to cheat the phone company's punitive long-distance tolls — the ISP came back online only after posting a 60,000 guilder (US\$35,000) cash bond. Just four years later, after XS4ALL had grown into one of the most successful ISPs in the lowlands, the state company bought out its former gadfly. Today XS4ALL is as independent as a subsidiary of a former government monopoly can be, but its members are not above sharing digs with their corporate parent, especially as the corporate parent is such a spendy sort of sugar-daddy. XS4ALL has taken over two storeys of the would-be switching centre with hackerish humour: the raised floors sport Perspex panels revealing neon lights and jumbles of entombed PC junk; there is a chill-out room for sysadmins who come on-site to run backups or swap drives; a poster listing the facility's regulations ends: "Rule 12: No sex in the data centre."

The mix of freewheeling hacker humour, deadly serious commitment to free speech and solid technological infrastructure made XS4ALL a natural choice to host a mirror copy of the Internet Archive (archive.org), a massive 'repository of all knowledge'. The archive's best-known feature is the Wayback Machine, an archive of most of the public pages on the World Wide Web that allows visitors to 'travel in time' and see what any given URL looked like on any given date since 1996. But it also serves as a repository for practically every public domain and Creative Commons-licensed digital document it can lay its hands on. It is the brainchild of philanthropist Brewster Kahle — co-creator of Wide Area Information Servers, or WAIS, one of the first Internet search engines — who wants it to provide "universal access to all knowledge". In a world of here today/gone tomorrow Web 2.0 companies willing to host your video, pictures or text, the archive stands out as a sober-sided grown-up with a commitment to long-term (infinite-term) hosting.

Inside the XS4ALL data centre, which is about the size of a football pitch, my hosts took me past aisle after aisle of roaring machines to the Internet Archive's European mirror.

"That's it, huh?"

Two racks, each the size of a modest refrigerator, each holding north of a petabyte's worth of information. These are the PetaBoxes, the Internet Archive's web-in-a-box systems. Designed as a shippable backup for every freely shareable file anyone cares to upload and hundreds of copies of everything else, too, they betray the archive's US origins in the strip of American-style electric outlets running down

one strut, a column of surprised clown-faces Fed-Exed from across the ocean. A couple of other things set them apart. Each rack draws 12 kilowatts, whereas a normal rack at the facility draws 4.5 kilowatts; the drive-housings are covered in a rather handsome fire-engine-red enamel. Apart from that, the PetaBoxes are just another pair of racks.

Yet housed in these machines are hundreds of copies of the web — every splenetic message-board thrash; every dry e-government document; every scientific paper; every pornographic ramble; every libel; every copyright infringement; every chunk of source code (for sufficiently large values of 'every', of course).

They have the elegant, explosive compactness of plutonium.

Something dawns on me: I ask my XS4ALL

tour guides, shouting over the jet-engine roar: "If there are all those copies of the web on the PetaBoxes, what's in all the other machines?"

"Oh, customer stuff.

Intranets. Databases. E-mail. Usenet." In other words, all the dynamic stuff, the private stuff, the dark web that is invisible to search engines, and all the processor power needed to serve it. All the reasons that Google can't exist just in a couple of red PetaBoxes.

"How does KPN feel about housing these two extraordinary boxes?"

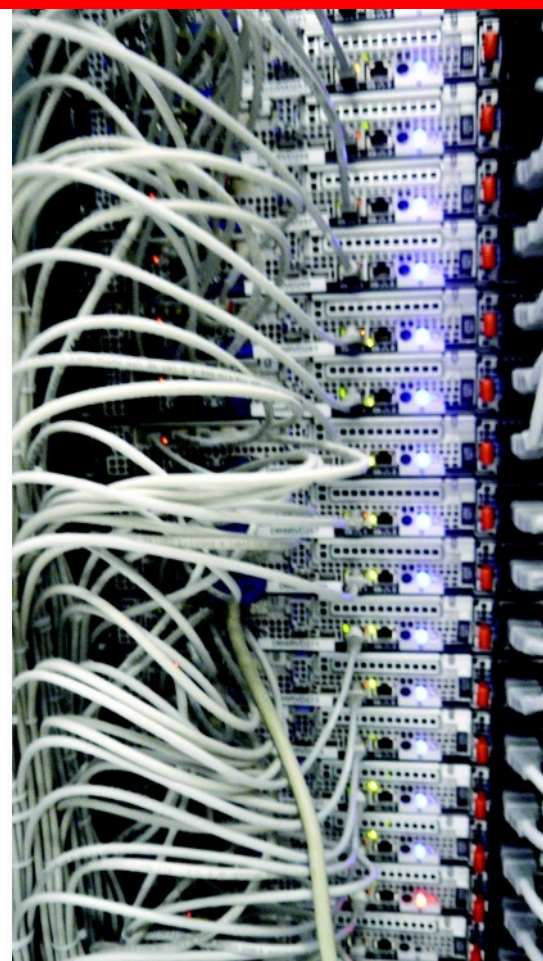
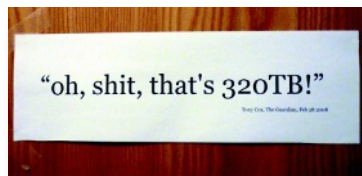
"Oh," they say, exchanging a mischievous glance, "I don't think they know we have them here."

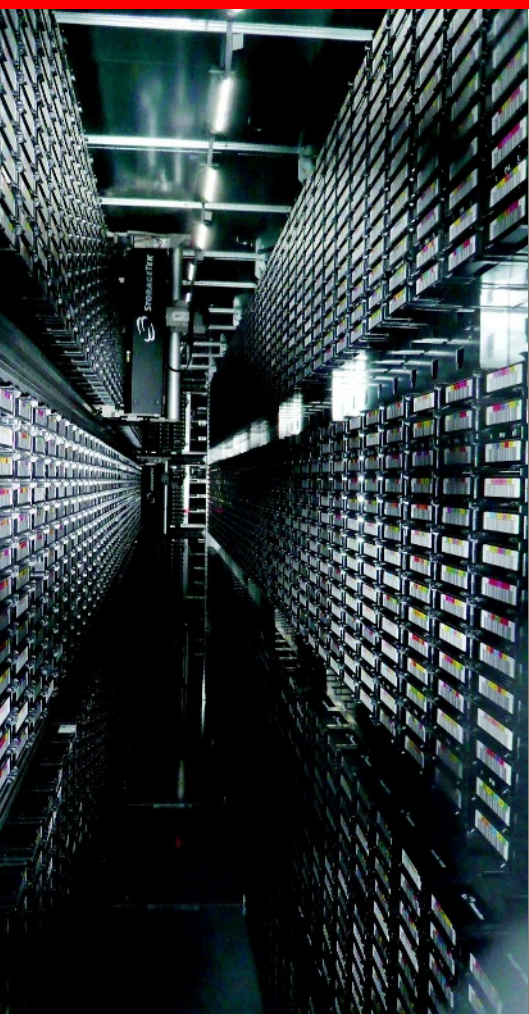
In a data centre such as this, a working approximation of 'all knowledge' can be slipped into the cracks like a 32-MB memory card jingling in my pocket.

600 million collisions a second

The archive has three real-time mirrors: the original in San Francisco's Presidio, just south of the Golden Gate, the XS4ALL mirror, and a third under the New Library of Alexandria in Egypt. A keen observer will note that these are variously placed on the San Andreas Fault, in a flood-zone, and in a country with a 27-years-and-running official 'state of emergency' that gives the government the power to arbitrarily restrict speech and publication. Someone needs to buy Kahle a giant cave in Switzerland. Like the one I'm off to now, which will be housing the data from the biggest experiments on the most powerful machine ever conceived.

Except it turns out that the data centre at CERN is less hall of the mountain king and more high-school gymnasium. The caverns measureless to man through which the





LHC runs are reserved for making the data. The systems storing them have much more humdrum quarters. The slight sense of anticlimax is emphasized by the unflappable calm of Tony Cass, CERN's leader of fabric infrastructure and operations; his data centre may be about to become the white-hot focus of the entire world's high-energy physics community, but Cass is surprisingly and perhaps bit disappointingly relaxed. Indeed, when we met just a few weeks before the LHC was about to see its first circulating beam, on 10 September, he was headed off on holiday.

Built in the 1970s to be the only data centre that CERN would ever need, Cass's current facility is now just a stopgap on the way to the construction of a bigger, faster centre that will absorb 15 petabytes a year of experimental data from the LHC. Although the rack after rack of systems in the current centre are nearly new, there are already plans to replace them. The basement is a graveyard of already-replaced generic PCs that are slowly being cleansed of data and shipped to bidders from the former Soviet Union.

The difference between Cass's challenges and Butcher's is a difference in the way that physics and biology work. At the Sanger, the charge-coupled devices (CCDs) in the sequencers can vomit out TIFF image files by the terabyte around the clock, but they are useless until processed, analysed and shrunk down to a far more manageable summary of what those vast image files actually meant. The original data are thrown away — the Sanger is confident that there will never be anything new to be learned from looking at the raw image files later. And there would be no way of keeping it except on tape, and tape, as Butcher will tell you, is slow, impractical and failure prone. As a former sysadmin myself, I can attest to the inherent tetchiness of tape. The Sanger reduces the images to more amenable data and then sends everything off to various mirror sites using a custom-made file-transfer protocol implemented over what's known as user datagram protocol (UDP); this allows the gene genies to saturate entire transoceanic links without having to wait for any of the finicky TCP handshaking and error-correction nonsense used for Internet traffic. It's slow compared with

CERN's approach — CERN leases its own fibre, at great expense — but it certainly beats the old-school open-source system used to glue together the Internet Archive's mirrors.

If only high-energy physics were so amenable to throwing stuff away, Cass's life would be a lot simpler. It's not. The meaning of a sequencer run is pretty straightforward, and won't change. The meaning of a particle collision is continuously reassessed based on new information about the instrument's performance. Physicists will want to reanalyse all the collisions expected from the LHC from the first to the last. Six hundred million events a second for year after year, analysed over and over again as the physicists' models become more refined. And that means a lot of storage — the kind of storage you can't load onto a reasonable quantity of spinning drives. The kind of storage you need to put on tape.

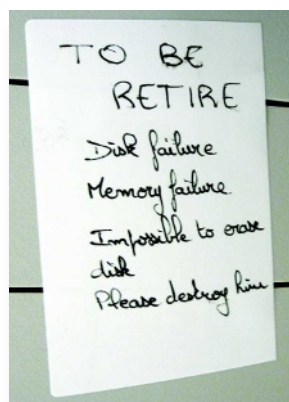
I am, admittedly, prone to swooning over a well-designed bit of IT kit, but I have never developed as deep and meaningful and instantaneous a relationship as the one I formed with the two tape-loading robots in the basement of the CERN data centres.

The Vader-black machines, one built by StorageTek, a subsidiary of Sun Microsystems, the other by IBM, are housed in square, meshed-in casings the size of small shipping containers.

From within them comes a continuous clacking noise like the rattling of steel polyhedral dice on a giant's Dungeons & Dragons table. I pressed my face against the mesh and peered in fascination at the robot arms zipping back and forth with tiny, precise movements, loading and unloading 500-GB tapes with the serene grace of Shaolin monks. Did I say tape is tetchy? I take it back. Tape is beautiful.

Each robot-librarian tends 5 PB of data. It will jump shortly to 10 PB each when the 500-GB tapes are switched to 1-TB models — an upgrade that will take a year of continuous load/read/load/write/discard operations, running in the interstices between the data centre's higher-priority tasks. When that is done, there should be 2-TB tapes to migrate to, bringing the two robots' total up to 40 PB. At least, that's what CERN hopes.

The tape libraries will allow the regular reassessment of the LHC data — unloading, reprocessing and reloading all the data on each of the tapes. A complete reprocessing will take a year, in part because, although it is higher priority than migrating the data to higher-density tapes, it still takes a backseat to the actual science —



A tape robot at CERN (top left); cooling devices at XS4ALL (bottom left); a label (above) for obsolete CERN computers; and rescue disks at CERN (facing page).

to jobs requested from anywhere in the world.

CERN embodies borderlessness. The Swiss–French border is a drainage ditch running to one side of the cafeteria; it was shifted a few metres to allow that excellent establishment to trade the finicky French health codes for the more laissez-fair Swiss jurisdiction. And in the data sphere it is utterly global.

Cass's operation is backstopped by ten 'Tier One' facilities around the world that replicate its tape library, and some hundreds of 'Tier Two' facilities that provide compute-power to operate on those data, all linked by dedicated, high-speed fibre, part of a global network that attempts to tie the world's high-energy physics institutions into a single, borderless facility. A researcher who logs into the CERN data centre runs code without worrying which processors execute it or which copy of the data it is run on. The birthplace of the web, which demolished borders for civilians, CERN is ushering in a borderless era for data-intensive science, an era in which US researchers run code on Iranian supercomputers and vice versa, without regard for their respective governments' sabre rattling. Cass wants to weld the world's physics computers into a single machine.

Sysadmin nightmares

At each data centre I asked the sysadmins for their worst fears. Universally, the answer was heat. Data centres are laid out in alternating cool and hot aisles, the cool looking at the front of the racks, the hot at the back. At CERN, they actually glass over the cool aisles to lower the cooling requirements, turning them into thrumming walk-in fridges lined with millions of tiny, twinkling lights.

If power is cut to the cooling system in one of these places, you've got minutes for a clean shutdown of the systems before their heat goes critical. XS4ALL has a particularly impressive cooling system, a loop that runs from the 5°C, 30-metre depths of nearby Lake Nieuwe Meer, warms to 16°C in the centre's exchangers, and then plunges back to the lake-bottom to be cooled again. The site manager Aryan Piets estimates that if it broke down and the emergency system didn't come on, the temperature in the centre would hit 42°C in ten minutes. No one could cleanly bring down all those machines in that time, and the dirtier the shutdown, the longer the subsequent start-up, with its rebuilding of databases and replacement of crashed components. Blow the shutdown and stuff starts to melt — or burn.

Data centres do face more exotic risks. Google once lost its transoceanic connectivity

because of shark bites. Butcher lives in fear of a Second World War fighter plane going astray from the airshows at nearby Duxford airfield and crashing into the Ice Box. At CERN they worry about people believing the worries that the Universe will wink out of existence when they fire up the LHC. But the real worry is power and its management. Data centres built in the giddy dotcom heyday assumed that racks would sport one processor core per unit and planned cooling and energy accordingly. But that is not the way the technology has gone. Computers have got faster not through faster cores, but through more of them. With 16 cores or more per unit, data centres around the world sit half-empty, unable to manage the power-appetites of a whole room's worth of 2008's ultra-dense computing. And everyone lives in fear of the electrical fault that sparks a blaze that wafts killing soot into the hungry ventilation intakes on the racks.

A big part of the problem — and possibly of its solution — is that most of a data centre's compute capacity is idle much of the time. No sysadmin would provision a data centre to run at capacity around the clock, lest it melt down (along with the sysadmin's career) the first time something really juicy increased the load. Yet whether a network card is saturated or idle, it still burns 100% of its energy draw. The same with video cards, power supplies, RAM and every other component except for some CPUs. So these idle systems whirl away, turning coal into electricity into heat that has to be cooled with coal turned into electricity turned into heat, and the planet warms and the bills soar. Every decibel of noise roaring through the centres is waste, energy pissed away for no benefit.

The people with the biggest data centres have the biggest problem — and the biggest resource to throw at it. Google buys its systems in enough bulk that it can lay down the law to component suppliers, demanding parts that draw power proportional to the amount of work that they are doing. Its holistic approach to the data centres, treating each one as a single PC, means that it can plan for idleness and peak load alike, and keep the energy bills under control. Everyone agrees that something like this is the way forward, that the future of data centres must be cooler, and quieter.

That said, a certain discomforting noise has its advantages. "I don't want it to ever get too comfortable in here," says Cass. "I like it that people access us remotely. It just doesn't scale, having every scientist drop in to run jobs here in the centre."

And if Google leads the way because it has to feed people's need for Paris Hilton searches and peeks at their own roof on Google Earth, that is

quite fitting. Whereas scientists unzip new genomes and summon new particles from the roiling vacuum with technologies beyond compare, the secret of data storage and processing is a lot simpler: commodity components. There is a huge ingenuity in how you use them, cool them, arrange them and keep them from melting, but the basic ingredients

of a petacentre are the ingredients of life on the net. Everything I've seen on these trips was basically made out of the same stuff I've got lying around the flat. Gene-sequencers use multi-megapixel CCDs — cheap and cheerful in this era of digital photography — to generate TIFFs that I could open with the open-source image-manipulation program that came with my free Ubuntu GNU/Linux operating system. The hard-drives in the server cases are the same cheap, high-capacity Seagates and Toshibas that I have in the little box I stuck under the stairs and wired up to my telly to store away a couple of terabytes of video, audio and games.

A decade ago, a firm's 'mainframe' was a powerful beast made from specialized components never seen outside the cool, dust-free environs of the data centre. Today, mainframe is more synonymous with the creaky old legacy system that no one can be bothered to shut down because it is running an obscure piece of accounting software that would be a pain to port to a modern system. The need for special hardware just isn't there any more. Even Google's 'energy-proportional' future is just an expansion of the power-management and heat-dissipation technology developed for laptops, and any gains achieved on the server side will also come to our desktops. I've got everything I need lying around the office to make my own petacentre — I just need more of it. And a much bigger fridge. Or a cool-bottomed lake.

That said, I don't have a tape robot.

But I really, really want one.

Cory Doctorow is a digital-rights activist, author and co-editor of *Boing Boing*, a blog. His most recent novel is *Little Brother*.

See Editorial, page 1.



C. DOCTOROW

"If the emergency system didn't come on the temperature would hit 42°C in ten minutes."



WIKIOMICS

Pioneering biologists are trying to use wiki-type web pages to manage and interpret data, reports **Mitch Waldrop**. But will the wider research community go along with the experiment?

Alexander Pico remembers just when the idea hit him. In January 2007, he and his boss, Bruce Conklin, were discussing how to push their software tool for visualizing intracellular signalling pathways to the next level of interactivity — when Pico blurted out, “What we really need is a wiki!”



Well, it was an original thought at the time, says Pico, a software engineer in Conklin's laboratory in the Gladstone Institute of Cardiovascular Disease at the University of California, San Francisco. In retrospect, it was one of those ideas that strikes everywhere at once. As soon as he and his colleagues started giving talks about ‘WikiPathways’, as they called their project, someone in the audience would invariably say, “Ah — we had the exact same idea.”

Scientist-edited interactive ‘wiki’-type websites have proliferated over the past year or so (see table), to the point where research-

ers have begun to joke about the new science of ‘wikiomics’. All the sites are modelled on the popular user-edited, online encyclopedia Wikipedia, and all aim to help biologists turn the data flooding into the large public gene and protein databases into useful knowledge.

The flood is going to rise even faster, says Amos Bairoch, executive director of the Swiss Institute for Bioinformatics in Geneva and creator of Swiss-Prot, a predecessor to the international protein sequence database UniProt: “As the price keeps going down, we’re reaching the point where every genome that can be sequenced, will be sequenced,” he says.

Ultimately, that could mean the genomes of most of Earth's 1.8 million named species, along with individual variants produced by projects such as the ‘1000 Genomes’ programme for humans. And there's all the rest of the quantifiable information about life on Earth — data on protein structure and function, biomolecular interactions, signalling

and metabolic pathways, and much more. The challenge is to make sense of the deluge.

Teams of scientist-annotators at the data repositories make valiant efforts to keep up, and bioinformatics programmers devise increasingly sophisticated annotation algorithms to help. Scientists write review articles and textbooks to help. But it's still not enough.

Hence the proliferation of wikis, which have the potential to vastly multiply the number of annotators and bring in the most interested expertise: “The best people to do annotation are the researchers in the laboratories, the people who are producing this knowledge in the first place,” says bioinformatician Barend Mons at the University of Rotterdam in the Netherlands. Mons is one of the prime movers behind WikiProfessional Life Sciences, a site that links publications on a given topic and enables users to add their own annotations.

But will the bench scientists participate?

"This business of trying to capture data from the community has been around ever since there have been biological databases," says Ewan Birney of the European Bioinformatics Institute in Hinxton, UK. And the efforts always seem to fizzle out. Founders enthusiastically put up a lot of information on the site, but the 'community' — either too busy or too secretive to cooperate — never materializes. So how do the wiki proponents know that this time around will be different?

They don't. "This is an experiment," says Pico, echoing just about everyone in the wiki movement. He is optimistic, however. This June he attended a workshop at the University of California, San Diego, on new communication channels in biology. "Many of the people had come to this from prior attempts," he says, "and were very sober about the challenges." From ensuring usability to ensuring users, these challenges go beyond the technical. As the developers of WikiPathways and several others have found, a truly cooperative web-based community requires a change in thinking — a shift in the way scientists work and in the way they get credit for that work.

Take but no give

Conklin's original idea for software to help biologists visualize and draw pathways grew from his research exploring how hormones and their receptors direct tissue development. Pathway diagrams are flow-chart representations of the interactions between genes, proteins or metabolites involved in a particular cellular function, such as the response to an external signal. They enable researchers to interpret the biochemical functions of individual molecules in the broader cell-biological context. One protein might have a very limited function, marking another protein for destruction, for example. But seeing its place in a pathway gives a clue to the physiological significance of that tiny action and offers clues to the functions of similar-looking proteins.

Better still, says Conklin, pathways help make sense of DNA microarray data on gene expression. If administering a drug enhances the expression of a set of genes all involved in the same pathway, say one causing cell death, then that's an important clue to what is going on.

So, back in 1999, Conklin's lab began to develop software that would make it easier to visualize and modify cellular signalling pathways. Known as the Gene Map Annotator and Pathway Profiler (GenMAPP), it offered free, downloadable software that could turn a database of interactions into a pathway diagram, and also enabled the user to add a new entry to the database simply by

sketching in a new reaction. GenMAPP also offered the capability to match microarray gene-expression data against an extensive library of known pathways and identify the most likely matches.

To get the library started, says Conklin, "I went to Amazon.com and bought \$900 worth of textbooks. Then my students and I flipped through and redrew the pathways we found there by hand, making electronic versions." They figured the tedium was worth it. Their library would grow fast, as soon as researchers who downloaded the drawing tool began uploading their own pathways.

But the team was overly optimistic. The GenMAPP drawing tool proved popular, and in the nine years since the launch, it's been downloaded 17,000 times. But when it came to giving back to the library — the rate wasn't so great. Only about 30% of the 557 pathways in the current GenMAPP library have come from outside the developers' own labs.

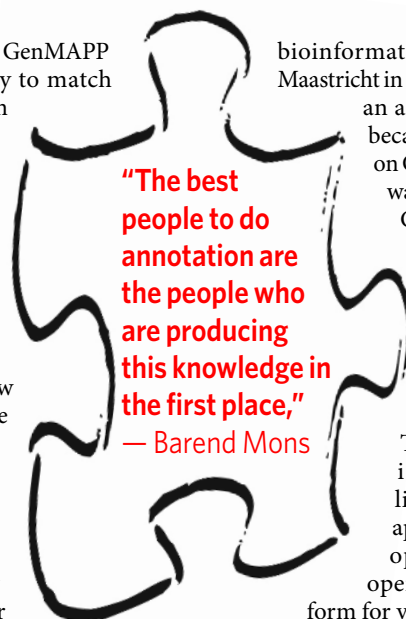
There were some enthusiasts. The group run by Chris Evelo, head of the department of

bioinformatics at the University of Maastricht in the Netherlands, was such an active contributor that it became a formal collaborator on GenMAPP in 2003. "But it was frustratingly slow," says Conklin. "We'd see publications with pathways created using our software, but half the time people wouldn't submit them back to us."

Make it easy

Two things broke the impasse, says Conklin. In 2005, the lab was approached by the developers of Cytoscape, an open-source software platform for very powerful, very high-end network analysis, much used in systems biology. They liked GenMAPP's layout, with its easy-to-use sketching capability, which they wanted to incorporate into Cytoscape, where the pathway drawings were abstract and mathematically elegant, but hard for the uninitiated to understand.

Conklin and his group were happy to oblige. "Cytoscape turned out to be supported by a very robust open-source community, which we didn't have," says Conklin. "Here were people



A SELECTION OF WIKI-STYLE INFORMATION COOPERATIVES

Name	Content
EcoliWiki http://ecoliwiki.net	The community-annotation component of EcoliHub, which integrates information from 19 websites relevant to <i>Escherichia coli</i> .
Gene Wiki http://en.wikipedia.org/wiki/Gene_Wiki	Not a site by itself, but an effort to create or update Wikipedia pages on some 9,000 human genes, using data and text from primary gene and protein databases
OpenWetWare http://openwetware.org	One of the oldest and largest scientist-edited sites, OpenWetWare has evolved into an active social network for biologists, hosting blogs, links to labs and special-interest groups and, of course, a long list of lab protocols
PDBWiki http://pdbwiki.org	A community-annotated knowledge base of biological molecular structures in the Protein Data Bank (PDB)
Proteopedia www.proteopedia.org	An interactive encyclopedia of three-dimensional structures of proteins, RNA, DNA and other molecules
Topsan www.topsan.org	The Open Protein Structure Annotation Network: a wiki devoted to protein three-dimensional structures, including some not yet deposited in the PDB
WikiGenes www.wikigenes.org	A specially designed wiki that tracks — and thus allows scientists to get credit for — every contribution that's made
WikiPathways www.wikipathways.org	A wiki devoted to the community curation and enhancement of biological pathways
Wiki Professional Life Sciences www.wikiprofessional.org	An attempt to create a 'Concept Web' by linking publications on a given 'concept' and enabling user annotation

coming out of the walls, offering us all kinds of software solutions.” The GenMAPP team became active participants — Conklin now sits on the Cytoscape board — and soon decided to revamp their own drawing tool entirely; the next GenMAPP release, due out in 2009, will essentially be a slightly specialized version of Cytoscape.

That involvement led to the second innovation, says Pico. “The Cytoscape team was using a wiki to coordinate their work,” he says. “And that was my first experience with the idea.” So he decided to install a wiki in Conklin’s lab for internal use. “These were mostly wet-lab biologists, and what impressed me was that even the least technically inclined people in the group picked it right up,” says Pico. “Even biologists who would never add to a website would add to the wiki — it was easy and fun.”

Sketching the idea

So the next big idea was almost inevitable — a public wiki interface for GenMAPP to make it easier for researchers to contribute their new pathways.

As inspiration hit on that January day, Pico sketched out his idea. It would need an online version of the GenMAPP drawing tool, instead of a separate piece of software to download, and a one-click submission of a finished pathway to the library instead of a separate uploading process. When he e-mailed Evelo with the idea, two of Evelo’s graduate students, Martijn van Iersel and Thomas Kelder, replied. Surprise — they’d had the same idea.

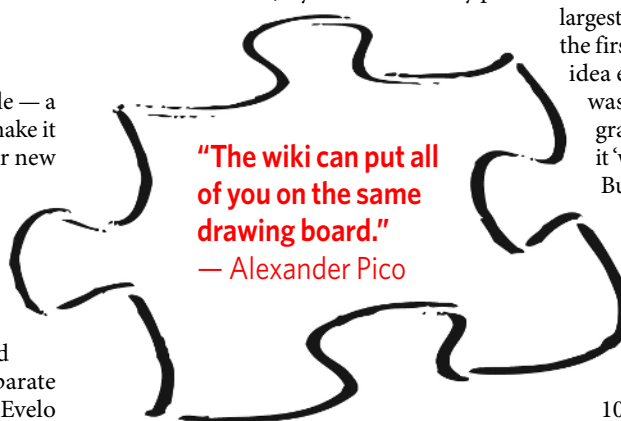
Kelder and van Iersel in Maastricht and Pico and Kristina Hanspers at the University of San Francisco became the design group for WikiPathways. A top priority was to make the site very easy for bench scientists to use. Like most of the other wiki-inspired biosites, WikiPathways does this by using the open-source MediaWiki software that underlies Wikipedia. As EcoliWiki creator James Hu of Texas A&M University in College Station puts it, “We didn’t want to ask young scientists who were already editing Wikipedia to learn a new interface.”

The WikiPathways team did need tools not available on Wikipedia itself. “We completely gutted the MediaWiki text-editing functionality and replaced it with new applets that would represent pathway information graphically,” says Pico. The diagram is linked behind the scenes to a structured database of biochemical interactions, he says, but the goal is to make drawing the pathway on screen as easy as drawing it on a napkin. “And then once you’re done, it’s immediately available to you — or to the world. You can e-mail the link and do

collaborative editing with biologists globally, which is impossible with GenMAPP or any other tool that’s on your personal machine,” says Pico. “The wiki can put all of you on the same drawing board.”

A prototype WikiPathways was up and running by spring 2007. By autumn the team felt confident enough to promote the site more widely. And in January 2008 they got their first pathway contributed by a researcher they didn’t know directly. “I consider that the birthday,” says Pico.

By mid-summer 2008, WikiPathways had some 350 registered users, of whom 50 or so had made changes to at least one pathway. “It’s already more contributors than we’d gotten over the past nine years,” says Pico. And for several weeks after July 2008, when they pub-



lished a description of WikiPathways in the journal *PLoS Biology* (see A. R. Pico *et al.* *PLoS Biol.* 6, e184; 2008), the average of one new pathway contributed per month jumped to a new pathway every other day. The hope is that at some point soon, says Pico, “we’ll reach a tipping point, a critical mass, where people from areas of biology we know nothing about will start participating in the whole cycle of revision and correction while involving us less and less — and it will become self-sustaining”.

Critical mass

WikiPathways is a stand-alone site, but a few of the new bio-wiki sites are tapped into Wikipedia directly. Earlier this summer, a team led by Andrew Su at the Genomics Institute of the Novartis Research Foundation in La Jolla, California, launched a software ‘robot’ that systematically goes through Wikipedia creating or amending entries for every human gene that has been studied to any significant degree — some 9,000 in all. The result is Gene Wiki: a collection of Wikipedia pages in a standard format, populated with an integrated suite of information culled from the National Center

for Biotechnology Information’s Entrez Gene, together with links to data repositories and publications, and to Wikipedia’s rich resource of pages on diseases and physiology. Gene Wiki entries are already showing up on the first page of Google search results for particular genes, says Su. “And our hope is that some number of readers will actually stay to make an edit,” he says. “It could be as trivial as fixing a typo, or as substantive as summarizing a new paper in the literature. But it will start a positive feedback loop by making the page that much more useful.”

Building critical mass — that’s the real challenge in the wiki game, as everyone is acutely aware. It’s also a mysterious process that requires timing and luck just as much as skill.

Wikipedia, for example, didn’t become the largest collaborative site on the planet by being the first. That honour goes to a programmers’ idea exchange called WikiWikiWeb, which was developed in 1995 by American programmer Ward Cunningham. (He named it ‘wiki’ after the Hawaiian word for ‘quick.’)

But Wikipedia, founded in 2001 by developers Jimmy Wales and Larry Sanger, was among the first to offer a free service — knowledge aggregation — that was useful to essentially any literate person on the planet. It proceeded to grow exponentially, to the point where it now claims more than

10 million articles in more than 250 languages — roughly a quarter of them in English. Wikipedia has also acquired the classic ‘long tail’ of contributors, with a comparative handful of people making lots of edits, and a multitude who make only a few.

The science wikis face a tougher challenge in building critical mass, if only because they’re aiming at a much smaller audience. One obvious strategy is to avoid fragmenting that audience. As Evelo points out, “biologists aren’t going to work on a dozen wikis to see which will survive”. They are going to want the various wikis to be interoperable and mutually supporting, so that the data they enter in one can be easily ported to another — or will even flow to all the appropriate sites automatically.

It should help that so many of the sites are based on the same MediaWiki software. That gives them the potential to act as one big open-source community, sharing code and improvements. And it’s not just potential, adds Pico: “We’ve been in close contact with Jim Hu and the Ecolihub folks about making our wikis interoperable.” Ecolihub is the ‘parent’ website of EcoliWiki, providing access to vast amounts of information on the bacterium *Escherichia coli*. Also critical to interoperability will be a standard language that can be understood by all

the databases. In the realm of pathways, says Pico, the closest to that right now is BioPAX, an XML-based standard for the exchange of pathway and interaction information. "We're planning on converting our system to it."

Interoperability is only part of the equation, however. Few scientists will contribute to these sites out of altruism. They need tangible incentives — starting with a real benefit to their day-to-day research.

Giving them that is definitely a work in progress, says van Iersel. The wiki architecture offers some possibilities. "For example, you can sign up to be e-mailed whenever a change is made to a page you're interested in," he says. So a researcher could immediately be alerted to any new findings in an area he or she is working on, not to mention the existence of potential collaborators (or rivals), without having to wait for a paper to come out.

There will always be some hypercompetitive fields in which people will keep their work under wraps for fear of getting scooped. But the hope is that for most researchers, the win-win dynamics of real-time data sharing will prevail. "Community annotation supports the natural process in which people form intellectual networks around topics," says Mons. "The system tells me, 'Hey — if you're interested in ABC, you'd better look at XYZ, as well.' And that will become part of the workflow of a scientist's life."

Due credit

Academic culture being what it is, however, the wiki sites will have to crack the credit-assignment problem, and provide some way for scientists' efforts there to be identified, recognized, cited and shown to funding agencies and tenure committees. Without a solution, says Hu, wiki-based community annotation will get nowhere. "Everybody gets excited by the idea," he says, "but then it always falls off the table, because it's not one of the things that pays the rent." Pico couldn't agree more. At the San Diego workshop, "we had break-out sessions, lunches, lots of brainstorming trying to think of metrics for scientists to quantify

their activity at these sites," he says.

Perhaps the most thoroughly worked out demonstration of how credit assignment could function in a wiki context is WikiGenes (not to be confused with Gene Wiki), created by Massachusetts Institute of Technology computer scientist Robert Hoffmann (see R. Hoffmann *Nature Genet.* **40**, 1047–1051; 2008). Like Wikipedia, the WikiGenes site consists of articles that are collaboratively written and edited by the users. Unlike Wikipedia, however, WikiGenes links every piece of text directly to its author. (In principle, a user could find that information on Wikipedia by tracking back through every previous version of an article, but in practice this rapidly becomes unworkable.) A single click leads to an automatically constructed page for that author, which lists all his or her contributions. Registered users have the option to do a one-click rating of each contribution, thus providing a fine-grained community peer review.

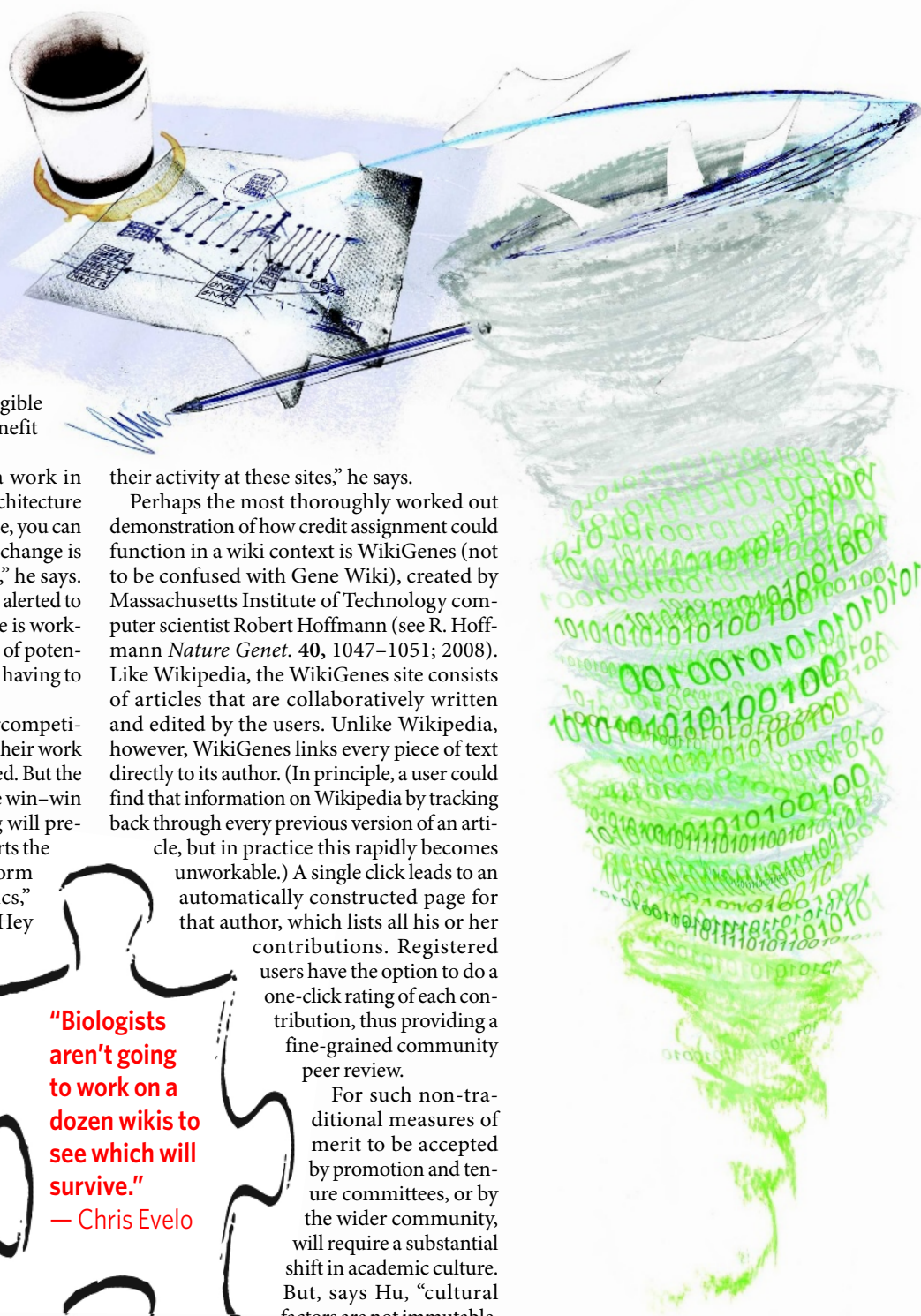
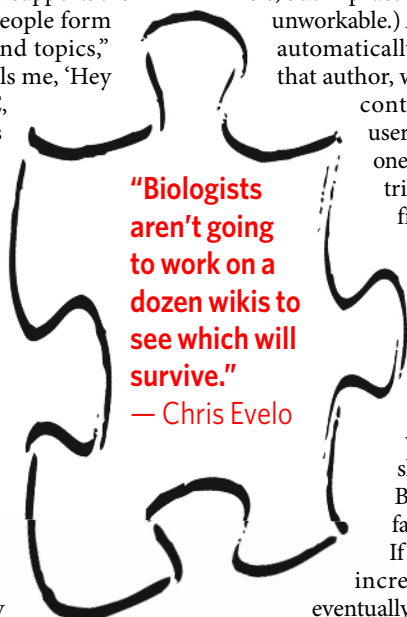
For such non-traditional measures of merit to be accepted by promotion and tenure committees, or by the wider community, will require a substantial shift in academic culture. But, says Hu, "cultural factors are not immutable.

If we can promote various incremental changes, then eventually this will take off."

In the meantime, the wikis still have a lot of challenges to face — not least the need to prove to funding agencies that they are worthy of long-term support.

And that is why it is all very much still an experiment. "Community intelligence is a new concept for biology — and in broader society — and we certainly don't claim to have the final

"Biologists aren't going to work on a dozen wikis to see which will survive."
— Chris Evelo



answer," says Su. Still, the more mechanisms for harnessing community intelligence, the better: "The community will essentially vote on which model will be the most useful, and the beauty is that they will vote with their participation," he says. "The only question is which model will resonate best."

Mitch Waldrop is Editorials and Features editor for Nature.

See also Editorial, page 1.

CORRESPONDENCE

Better writing and more space needed online

SIR — The World-Wide Web is remarkable as a vehicle for communicating scientific discoveries. Online journals unite distant researchers and inspire worldwide collaborations. However, despite these advantages, there is a growing risk that papers published today are less successful in meeting their objectives than in the past.

To ensure clear communication, most journals encourage authors to write for a broad audience. But most published papers still compress too much information into uncomfortably short articles, leading to convoluted sentences, specialized terminology and a proliferation of abbreviations. Errors in grammatical style result in impenetrable and ambiguous texts that seriously undermine the scientific literature. This need not be the case.

Electronic publishing could offer authors limitless space to explain their ideas and discuss their new findings. Surprisingly, though, online manuscripts are often bound by the same space constraints as print manuscripts.

Authors are instructed to conform to print-journal guidelines, leading many to redirect essential material to online Supplementary Information. The recent explosion in Supplementary Information is problematic: it seems to have no standard format among different journals, and there is a common misperception that data in Supplementary Information have escaped peer review. It can be a nuisance for readers too. For example, if they want to peruse articles away from their computers and haven't downloaded the related Supplementary Information, it may be impossible for them to understand or fully evaluate the papers' merits.

The scientific article in 2008 is on the cusp of change, with one foot in the past and one in the

future. Science journals should shed the constraints of the old media and exploit the advantages of the new, to offer readers easy and enjoyable access to the scientific literature.

Even if journals are successful at reinventing themselves, it won't be adequate unless the quality of writing in scientific manuscripts improves. Paradoxically, the deterioration in science writing seems to coincide with the swell in e-publications — at a time when the need to communicate advances in science is more urgent than ever. The quality of writing needs to match the power of today's e-publishing technology.

Linda Cooper Redpath Museum, Faculty of Science, McGill University, Montreal, Quebec H3A 2K6 Canada
e-mail: linda.cooper@mcgill.ca

Languages: Catalan speakers learn a wider range

SIR — Jose M. Rojo claims, in his Correspondence 'Schools in a third of Spain teach only in minority languages' (*Nature* **454**, 575; 2008), that public education is not available in Spanish in schools in Catalonia, Mallorca and Valencia. However, in Catalonia, the Spanish-language skills of schoolchildren completing their education are equivalent to those of children across Spain.

The Programme for International Student Assessment (www.pisa.oecd.org) indicates that the learning capacities of Catalan and Spanish schoolchildren in science and mathematics are not dependent on whether they receive a bilingual education. This conclusion flies in the face of the manifesto mentioned in Rojo's letter, which seeks to enforce a Spanish rather than bilingual education, and to relegate Basque, Catalan and Galician to a linguistic ghetto.

A recent study shows that, in most Spanish regions, between half and two-thirds of



the population does not know a foreign language (F. Alvira Martín and J. García López *Cuad. Inform. Econ.* **205**, 119–138; 2008; <http://tinyurl.com/64ngkh>). But in Catalonia and the Balearic Islands, where most of the population understands both Catalan and Spanish, about three-quarters of the population can also speak a foreign language. It might be in the better interests of Spain and science to improve the present knowledge of foreign languages and encourage an effective multilingual education, rather than striving to enforce monolingual Spanish education.

Antoni Rosell-Melé Institute of Environmental Science and Technology (ICTA), Universitat Autònoma de Barcelona (UAB), Edifici Cn – Campus UAB, 08193 Bellaterra, Catalonia, Spain
e-mail: antoni.rosell@uab.cat

Languages: Spain's minority-language speakers are bilingual

SIR — In his Correspondence 'Schools in a third of Spain teach only in minority languages' (*Nature* **454**, 575; 2008), Jose M. Rojo complained about the impossibility of studying in Spanish in one-third of the public schools in Spain. This is, at best, misleading. The Catalan schooling system, for example, does indeed promote the use of Catalan, but native Catalan students are as fluent in Spanish as their monolingual counterparts. The political manifesto Rojo cites to

emphasize his point is riddled with contradictions, is not endorsed by any linguists and does not belong in the pages of *Nature*.

Jesús Purroy Scientific Department, Parc Científic de Barcelona, Baldiri Reixac 10, 08028 Barcelona, Catalonia, Spain
e-mail: jpurroy@pcb.ub.cat

Readers are welcome to comment at <http://tinyurl.com/5e6ltj>

Religion: science is partially based on faith

SIR — Andrew Brown's Obituary of John Templeton (*Nature* **454**, 290; 2008) and your Editorial ('Templeton's legacy' *Nature* **454**, 253–254; 2008) both touch upon the philanthropist's interest in science and faith. Some might argue that science and faith should be kept separate, although others have no problem in reconciling the two. I am reminded of the different perspective on this eternal debate that is offered in astrophysicist Carl Sagan's science-fiction novel *Contact* (Orbit, 1985) — though not in the film of the same name, which is only very loosely based on the book.

Contact recounts an astronomer's successful search for alien intelligence. It also has a subplot that science and religion are, in fact, closer than the two camps imagine. Scientists' use of the scientific method pragmatically includes faith. A scientist must first conceive

CORRESPONDENCE

Better writing and more space needed online

SIR — The World-Wide Web is remarkable as a vehicle for communicating scientific discoveries. Online journals unite distant researchers and inspire worldwide collaborations. However, despite these advantages, there is a growing risk that papers published today are less successful in meeting their objectives than in the past.

To ensure clear communication, most journals encourage authors to write for a broad audience. But most published papers still compress too much information into uncomfortably short articles, leading to convoluted sentences, specialized terminology and a proliferation of abbreviations. Errors in grammatical style result in impenetrable and ambiguous texts that seriously undermine the scientific literature. This need not be the case.

Electronic publishing could offer authors limitless space to explain their ideas and discuss their new findings. Surprisingly, though, online manuscripts are often bound by the same space constraints as print manuscripts.

Authors are instructed to conform to print-journal guidelines, leading many to redirect essential material to online Supplementary Information. The recent explosion in Supplementary Information is problematic: it seems to have no standard format among different journals, and there is a common misperception that data in Supplementary Information have escaped peer review. It can be a nuisance for readers too. For example, if they want to peruse articles away from their computers and haven't downloaded the related Supplementary Information, it may be impossible for them to understand or fully evaluate the papers' merits.

The scientific article in 2008 is on the cusp of change, with one foot in the past and one in the

future. Science journals should shed the constraints of the old media and exploit the advantages of the new, to offer readers easy and enjoyable access to the scientific literature.

Even if journals are successful at reinventing themselves, it won't be adequate unless the quality of writing in scientific manuscripts improves. Paradoxically, the deterioration in science writing seems to coincide with the swell in e-publications — at a time when the need to communicate advances in science is more urgent than ever. The quality of writing needs to match the power of today's e-publishing technology.

Linda Cooper Redpath Museum, Faculty of Science, McGill University, Montreal, Quebec H3A 2K6 Canada
e-mail: linda.cooper@mcgill.ca

Languages: Catalan speakers learn a wider range

SIR — Jose M. Rojo claims, in his Correspondence 'Schools in a third of Spain teach only in minority languages' (*Nature* **454**, 575; 2008), that public education is not available in Spanish in schools in Catalonia, Mallorca and Valencia. However, in Catalonia, the Spanish-language skills of schoolchildren completing their education are equivalent to those of children across Spain.

The Programme for International Student Assessment (www.pisa.oecd.org) indicates that the learning capacities of Catalan and Spanish schoolchildren in science and mathematics are not dependent on whether they receive a bilingual education. This conclusion flies in the face of the manifesto mentioned in Rojo's letter, which seeks to enforce a Spanish rather than bilingual education, and to relegate Basque, Catalan and Galician to a linguistic ghetto.

A recent study shows that, in most Spanish regions, between half and two-thirds of



the population does not know a foreign language (F. Alvira Martín and J. García López *Cuad. Inform. Econ.* **205**, 119–138; 2008; <http://tinyurl.com/64ngkh>). But in Catalonia and the Balearic Islands, where most of the population understands both Catalan and Spanish, about three-quarters of the population can also speak a foreign language. It might be in the better interests of Spain and science to improve the present knowledge of foreign languages and encourage an effective multilingual education, rather than striving to enforce monolingual Spanish education.

Antoni Rosell-Melé Institute of Environmental Science and Technology (ICTA), Universitat Autònoma de Barcelona (UAB), Edifici Cn – Campus UAB, 08193 Bellaterra, Catalonia, Spain
e-mail: antoni.rosell@uab.cat

Languages: Spain's minority-language speakers are bilingual

SIR — In his Correspondence 'Schools in a third of Spain teach only in minority languages' (*Nature* **454**, 575; 2008), Jose M. Rojo complained about the impossibility of studying in Spanish in one-third of the public schools in Spain. This is, at best, misleading. The Catalan schooling system, for example, does indeed promote the use of Catalan, but native Catalan students are as fluent in Spanish as their monolingual counterparts. The political manifesto Rojo cites to

emphasize his point is riddled with contradictions, is not endorsed by any linguists and does not belong in the pages of *Nature*.

Jesús Purroy Scientific Department, Parc Científic de Barcelona, Baldiri Reixac 10, 08028 Barcelona, Catalonia, Spain
e-mail: jpurroy@pcb.ub.cat

Readers are welcome to comment at <http://tinyurl.com/5e6ltj>

Religion: science is partially based on faith

SIR — Andrew Brown's Obituary of John Templeton (*Nature* **454**, 290; 2008) and your Editorial ('Templeton's legacy' *Nature* **454**, 253–254; 2008) both touch upon the philanthropist's interest in science and faith. Some might argue that science and faith should be kept separate, although others have no problem in reconciling the two. I am reminded of the different perspective on this eternal debate that is offered in astrophysicist Carl Sagan's science-fiction novel *Contact* (Orbit, 1985) — though not in the film of the same name, which is only very loosely based on the book.

Contact recounts an astronomer's successful search for alien intelligence. It also has a subplot that science and religion are, in fact, closer than the two camps imagine. Scientists' use of the scientific method pragmatically includes faith. A scientist must first conceive

"These women were proof that 'people power' is capable of great things." Sue Nelson, page 36

the idea for an experiment, and then — on the basis merely of the hopeful presumption of its possible outcome — invest time and resources in funding and executing it in the anticipation of a meaningful result.

Work supported by the Templeton Foundation that investigates the relationship between science and faith could help to improve science communication and to address science-and-society issues. So let's hope that Templeton's son has the same penchant for meaningfully verifiable results as his dad.

Jonathan Cowie Thurnby Lodge,
Leicester LE5 2WG, UK
<http://www.science-com.concatenation.org>

Vavilov's vision for genetics was among Stalin's many victims

SIR — Jan Witkowski's review of Peter Pringle's fascinating and timely book on the famous geneticist Nikolai Vavilov ('Stalin's war on genetic science' *Nature* **454**, 577–579; 2008) is informative, but contains some oversimplifications and inaccuracies.

The review pays little credit to Vavilov as a unique theoretician, not just a practitioner of applied science. His intentions were not simply to feed the people or to cultivate sturdy mountain plants. His was a grander vision, worthy of his teacher William Bateson: to bring modern genetics into agriculture, to collect global data on his famous "homological series [parallelisms] in hereditary variation" and cultural plant centres of origin, and to compile global gene collections.

Because of the fraudulent geneticist Trofim Lysenko, a giant system of data falsification developed in the USSR. The subjects were forced to praise the emperor's new clothes where there were none. The relationship between Lysenko and Vavilov was indeed complicated: Vavilov first

promoted Lysenko's vernalization experiments and his career. The totalitarian and unpredictable nature of Stalin's regime not only prevented free criticism of Lysenko's data and his primitive 'Soviet genetics', it also led to the destruction of critics and opponents. Biology was a front line in the ideological war waged against Western ('bourgeois') science.

To call Stalin's agricultural collectivization policy a "consolidation of land and labour" is an awful understatement: an estimated 10 million productive peasants and their families were exiled or imprisoned from 1929–1933. Stalin was hardly "desperate to feed thousands of citizens dying of starvation" when these were the same people he starved and murdered while sending Russian grain abroad.

No free discussion about "the best data available" was possible for scientists in 1930s Russia. Saying that "even now, politics continues to trump good science" should not be taken as equating murderous dictators with democratic governments.

Victor Fet Department of Biological Sciences, Marshall University, Huntington, West Virginia 25755, USA
e-mail: fet@marshall.edu
Michael D. Golubovsky Department of Molecular and Cell Biology, University of California–Berkeley, Berkeley, California 94720, USA

Message from the heavens may be that there is no message

SIR — In his Opinion piece 'Message from the heavens' (*Nature* **453**, 1185; 2008), Martin Kemp tries to discern the meaning behind Maurizio Cattelan's shocking sculpture of Pope John Paul II felled by a meteorite. Although acknowledging that this sculpture has much in common with Marcel Duchamp's anti-art, he proceeds to provide a range of possible interpretations that include seeing it as an allegory of

the conflict between Darwinists and those with spiritual beliefs.

As the artist himself has chosen to remain silent on the topic (maybe wisely so), perhaps one should view this kind of art as a successful attempt simply to attract attention. Attention is such an important resource that people (scientists included) are willing to forsake financial gain to secure it. From this perspective, Cattelan's work fits an artistic tradition exemplified by people like Duchamp and Andy Warhol: masters at putting together pieces whose sole purpose was to grab our attention.

In a world increasingly awash with 'content creators' and the all-too-human limited attention we can devote to them, I see this work as a superb attempt to generate novelty and shock — to make us sit up and concentrate, even if only fleetingly.

Bernardo A. Huberman Social Computing Lab, HP Laboratories, 1501 Page Mill Road MS 1139, Palo Alto, California 94304, USA
e-mail: bernardo.huberman@hp.com

Senior staff of Mexican institute speak up

SIR — We find that your News story 'Scientists rally to Mexican researchers' plea' (*Nature* **454**, 143; 2008) is unjustifiably biased in favour of Harold Kroto and the research group of the Terrones brothers whom he defends.

Our institute for scientific and technological research, IPICYT, is one of 27 nationwide research centres coordinated by Mexico's national council of science and technology (CONACYT). This relies on long-established mechanisms for selecting the best researchers and directors. The present director of IPICYT, David Rios Jara, is supported by all the other CONACYT directors and by different Mexican academic organizations in his stand on the Terrones brothers affair.

IPICYT comprises five highly

successful multidisciplinary divisions and the national supercomputing centre, which between them operate four prestigious graduate programmes. The advanced-materials department (AMD) where the Terrones work represents about 20% of IPICYT's academic output.

The conflict involving the Terrones brothers attracted international attention because of their scientific reputation and connections with foreign scientists. These would not have been possible without the exceptional treatment and financial support they received at the hands of the former and current IPICYT directors. The AMD researchers, students, postdocs and technicians continue to work normally, despite the Terrones' claim that their group is being harassed and thwarted. The group remains the most well supported at IPICYT.

In relieving Humberto Terrones of his administrative duties, after more than seven years as AMD's head, Rios Jara was not persecuting him but was simply complying with the recommendation by the last external evaluating committee and the CONACYT governing board. One intention in removing these duties was to improve relations between the Terrones group and the rest of the AMD.

Mexican science is definitely not under threat, neither will it be affected by changing a single division head of a CONACYT centre. Indeed, the new measures enable the Terrones to enjoy more time on their research, which should help to boost their scientific output.

Carlos Barajas-López and senior staff members* Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), Camino a la Presa San José 2055, Col. Lomas 4a. Secc. SLP, CP78216, México
e-mail: cbarajas@ipicyt.edu.mx

*See supplementary information for full author list

Contributions may be submitted to correspondence@nature.com.

"These women were proof that 'people power' is capable of great things." Sue Nelson, page 36

the idea for an experiment, and then — on the basis merely of the hopeful presumption of its possible outcome — invest time and resources in funding and executing it in the anticipation of a meaningful result.

Work supported by the Templeton Foundation that investigates the relationship between science and faith could help to improve science communication and to address science-and-society issues. So let's hope that Templeton's son has the same penchant for meaningfully verifiable results as his dad.

Jonathan Cowie Thurnby Lodge,
Leicester LE5 2WG, UK
<http://www.science-com.concatenation.org>

Vavilov's vision for genetics was among Stalin's many victims

SIR — Jan Witkowski's review of Peter Pringle's fascinating and timely book on the famous geneticist Nikolai Vavilov ('Stalin's war on genetic science' *Nature* **454**, 577–579; 2008) is informative, but contains some oversimplifications and inaccuracies.

The review pays little credit to Vavilov as a unique theoretician, not just a practitioner of applied science. His intentions were not simply to feed the people or to cultivate sturdy mountain plants. His was a grander vision, worthy of his teacher William Bateson: to bring modern genetics into agriculture, to collect global data on his famous "homological series [parallelisms] in hereditary variation" and cultural plant centres of origin, and to compile global gene collections.

Because of the fraudulent geneticist Trofim Lysenko, a giant system of data falsification developed in the USSR. The subjects were forced to praise the emperor's new clothes where there were none. The relationship between Lysenko and Vavilov was indeed complicated: Vavilov first

promoted Lysenko's vernalization experiments and his career. The totalitarian and unpredictable nature of Stalin's regime not only prevented free criticism of Lysenko's data and his primitive 'Soviet genetics', it also led to the destruction of critics and opponents. Biology was a front line in the ideological war waged against Western ('bourgeois') science.

To call Stalin's agricultural collectivization policy a "consolidation of land and labour" is an awful understatement: an estimated 10 million productive peasants and their families were exiled or imprisoned from 1929–1933. Stalin was hardly "desperate to feed thousands of citizens dying of starvation" when these were the same people he starved and murdered while sending Russian grain abroad.

No free discussion about "the best data available" was possible for scientists in 1930s Russia. Saying that "even now, politics continues to trump good science" should not be taken as equating murderous dictators with democratic governments.

Victor Fet Department of Biological Sciences, Marshall University, Huntington, West Virginia 25755, USA
e-mail: fet@marshall.edu
Michael D. Golubovsky Department of Molecular and Cell Biology, University of California–Berkeley, Berkeley, California 94720, USA

Message from the heavens may be that there is no message

SIR — In his Opinion piece 'Message from the heavens' (*Nature* **453**, 1185; 2008), Martin Kemp tries to discern the meaning behind Maurizio Cattelan's shocking sculpture of Pope John Paul II felled by a meteorite. Although acknowledging that this sculpture has much in common with Marcel Duchamp's anti-art, he proceeds to provide a range of possible interpretations that include seeing it as an allegory of

the conflict between Darwinists and those with spiritual beliefs.

As the artist himself has chosen to remain silent on the topic (maybe wisely so), perhaps one should view this kind of art as a successful attempt simply to attract attention. Attention is such an important resource that people (scientists included) are willing to forsake financial gain to secure it. From this perspective, Cattelan's work fits an artistic tradition exemplified by people like Duchamp and Andy Warhol: masters at putting together pieces whose sole purpose was to grab our attention.

In a world increasingly awash with 'content creators' and the all-too-human limited attention we can devote to them, I see this work as a superb attempt to generate novelty and shock — to make us sit up and concentrate, even if only fleetingly.

Bernardo A. Huberman Social Computing Lab, HP Laboratories, 1501 Page Mill Road MS 1139, Palo Alto, California 94304, USA
e-mail: bernardo.huberman@hp.com

Senior staff of Mexican institute speak up

SIR — We find that your News story 'Scientists rally to Mexican researchers' plea' (*Nature* **454**, 143; 2008) is unjustifiably biased in favour of Harold Kroto and the research group of the Terrones brothers whom he defends.

Our institute for scientific and technological research, IPICYT, is one of 27 nationwide research centres coordinated by Mexico's national council of science and technology (CONACYT). This relies on long-established mechanisms for selecting the best researchers and directors. The present director of IPICYT, David Rios Jara, is supported by all the other CONACYT directors and by different Mexican academic organizations in his stand on the Terrones brothers affair.

IPICYT comprises five highly

successful multidisciplinary divisions and the national supercomputing centre, which between them operate four prestigious graduate programmes. The advanced-materials department (AMD) where the Terrones work represents about 20% of IPICYT's academic output.

The conflict involving the Terrones brothers attracted international attention because of their scientific reputation and connections with foreign scientists. These would not have been possible without the exceptional treatment and financial support they received at the hands of the former and current IPICYT directors. The AMD researchers, students, postdocs and technicians continue to work normally, despite the Terrones' claim that their group is being harassed and thwarted. The group remains the most well supported at IPICYT.

In relieving Humberto Terrones of his administrative duties, after more than seven years as AMD's head, Rios Jara was not persecuting him but was simply complying with the recommendation by the last external evaluating committee and the CONACYT governing board. One intention in removing these duties was to improve relations between the Terrones group and the rest of the AMD.

Mexican science is definitely not under threat, neither will it be affected by changing a single division head of a CONACYT centre. Indeed, the new measures enable the Terrones to enjoy more time on their research, which should help to boost their scientific output.

Carlos Barajas-López and senior staff members* Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), Camino a la Presa San José 2055, Col. Lomas 4a. Secc. SLP, CP78216, México
e-mail: cbarajas@ipicyt.edu.mx

*See supplementary information for full author list

Contributions may be submitted to correspondence@nature.com.

COMMENTARY

How do your data grow?

Scientists need to ensure that their results will be managed for the long haul.

Maintaining data takes big organization, says **Clifford Lynch**.

Data can be 'big' in different ways. National and international projects such as the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva in Switzerland, or the Large Synoptic Survey Telescope planned for northern Chile, are frequently cited for the way they will challenge the state of the art in computation, networking and data storage. But research data can also be big by being of lasting significance — a clinical-trial result, or the observation of a unique event. Data can be big because of descriptive challenges that may require context such as the experimental set-up. Because digital data are so easily shared and replicated and so recombinable, they present tremendous reuse opportunities, accelerating investigations already under way and taking advantage of past investments in science.

To enable reuse, data must be well preserved. In some cases the effects of data loss are economic, because experiments have to be re-run. In other cases, data loss represents an opportunity lost forever. Funders now rightly view data as assets that they are underwriting and so seek the greatest pay-off for their investments. They demand that researchers and host institutions document and implement data-management and data-sharing plans that address the full life cycle of data — including what happens after a grant finishes. Host universities thus find themselves with legal and ethical obligations to provide a legacy of faculty data. Publishers must also identify the most effective ways to connect publications with data and preserve the scientific record.

Developing infrastructure

Managing the life cycle of scientific data presents many challenges. These include deciding responsibilities, funding, resource allocation, what data should be kept and for how long.

In a sense, landmark international projects like the LHC are the least problematic: the costs of data management are explicit in the budget and tend to be dominated by technology expenses that decline over time. These projects also include dedicated personnel; and, although the volume of data is often vast, the streams fit within well defined descriptive schemes.

But science's reliance on digital data extends far beyond these international projects. Funding programmes in Europe and the United States, for



example, have invested substantially in common infrastructure for a more systematic reliance on data, networks and computation. And there are vast numbers of scientific research projects producing at most a few terabytes per year of big data, or data that can be aggregated into a big-data resource. Funding, support expertise and structuring the data for long-term management can be problematic for these projects. This has been shown in recent years by studies of faculty information management needs through a wide range of academic disciplines^{1,2}.

The challenges here are great, and will only be solved by focused effort and collaboration between funders, institutions and scientists.

Community standards for data description and exchange are crucial. These facilitate data reuse by making it easier to import, export, compare, combine and understand data. Standards also eliminate the need for each data creator to develop unique descriptive practices. They open the door to development of disciplinary repositories for specific classes of data and specialized software management tools. GenBank, the US National Institutes of Health (NIH) genetic sequence database, and the US National Virtual Observatory are good examples of what is possible here. In 2007, the US National Science Foundation, recognizing the importance of such standards, established the Community Based Data Interoperability Networks (INTEROP) funding programme for the development of tools, standards and data management best practices within specific disciplinary communities. INTEROP should make its first awards this autumn. Although many classes of scientific data aren't ready, or aren't appropriate, for standardization, well chosen investments in standardization show a consistently high pay-off³.

At the start of the data life cycle, individual scientists will have primary responsibility for stewardship. But longer term, data preservation can only be done by institutions. If data are to be consolidated or shared on a frequent basis, there is a lot to be said for moving to institutional control sooner rather than later. Scientists are not necessarily good data managers and can more fruitfully spend their time doing science. Moreover, it is unfair and unreasonable — and increasingly ineffective — to assign long-term

information management tasks to a rotating staff of students and postdocs. Indeed, as specific data sets become distant from current research activities, stewardship can become a tax on scientific productivity.

Scientists need to act responsibly during their stewardship. This includes working through and honouring disciplinary standards. It also includes defining and recording appropriate metadata — such as experimental parameters and set-up — to allow for data interpretation. This is best done when the data are captured. Indeed, descriptive metadata are often integrated within the experimental design. Description includes tracing provenance — where the data came from, how they were derived, their dependence on other data and all changes made since their capture. Proper stewardship requires documenting the storage formats. These may be community standards, or they may be locally defined and often tied to locally developed software. It is desirable to keep versions of such software along with the data sets.

If data cannot survive in the short term, it is pointless to talk about long-term use. In a high-threat environment such as a major university's network, machines will often be compromised if updates aren't applied; this can mean data

destruction or corruption. Disasters such as Hurricane Katrina, which destroyed labs and computing facilities, are important reminders that data need to be backed up

frequently and comprehensively in diverse and distant locations. Appropriate use of IT services such as secure storage or hosting from the host institution may be valuable. In the longer term, digital data is at risk from various forms of technological obsolescence (particularly if locally held removable storage media are used). There is a need for new institutional services that can help with all these needs, handling traditional IT issues and information-management issues more familiar to librarians and archivists.

At some point, the primary copy needs to migrate to an institutional service. Today, these services are sparse. In the United Kingdom there are data services associated with several of the science-funding councils. Both NASA and the European Space Agency have planetary

"The best stewardship of data will come from engagement with preservation institutions."



D. ALLISON

science archives into which they place mission data. And in the United States, for example, there are also other focused archives connected to some disciplines, including the collections at the NIH's National Center for Biotechnology Information, the social science archive of the Inter-University Consortium for Political and Social Research based at the University of Michigan in Ann Arbor, and the Protein Data Bank, which holds structural data for proteins and nucleic acids. These are somewhat mature and have relatively stable funding.

New disciplinary repositories are also springing up, and some universities are setting up broad-based multidisciplinary repository services, usually working through the campus research library, to manage their faculties' research data. The National Science Foundation is preparing to make its first awards under an Office of Cyberinfrastructure programme called Datanet that will invest around US\$100 million over the next five years for building data-stewardship capabilities; the grants will go to large university-led consortia. There are possible roles here for publishers and scholarly societies, but at present it seems as if in most disciplines, leadership will fall to stewardship services

run by universities and government agencies.

These newer institutionalized data stewardship services — whether structured along university or disciplinary lines — are still immature. The handing over of data for deposit is not simple or well defined, and necessary community standards are lacking. Funding models are sketchy. Although stewardship needs to be funded, funding agencies are not eager to pay. Educational institutions are equally reluctant to make open-ended commitments. Perhaps, ultimately, this can be factored into overhead cost negotiations. Effective structures are needed to manage limited resources; not everything can be preserved forever, and we need methods for prioritization.

Ultimately, the best stewardship of data will come from disciplinary engagement with preservation institutions. General-purpose data management as provided by universities through their research libraries will have its limits. Where there is no natural locus of disciplinary stewardship, universities will need to establish consortia to enable disciplines to create and sustain such engagement⁴.

The time is right for scientists to take stock of the institutionalized data services that are

available or under development, to understand how these institutions are governed and financed, and to make choices about the best strategies for their disciplines. Can a discipline-oriented solution work? If a university-based system seems more practical, what can be done to expedite the move to university consortia strategies? As the volume of data, and the need to manage it grows, disciplinary consensus leadership will be very powerful factors in addressing the challenges ahead. ■

Clifford Lynch is the executive director of the Coalition for Networked Information, 21 Dupont Circle, Washington DC 20036, USA, and an adjunct professor at the School of Information, University of California, Berkeley, California, 94720-4600, USA.
e-mail: cliff@cni.org

1. www.lib.umn.edu/about/scieval/documents.html
2. www.library.ucsb.edu/informatics/documents.html
3. www.ctwatch.org/quarterly/articles/2005/02/scientific-data-management/
4. ARL Workshop on New Collaborative Relationships Report to the National Science Foundation. *To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering* (2006).

See Editorial, page 1.

Join the discussion at <http://tinyurl.com/6eedyu>.

BOOKS & ARTS

Distilling meaning from data

Buried in vast streams of data are clues to new science. But we may need to craft new lenses to see them, explain **Felice Frankel** and **Rosalind Reid**.

It is a breathtaking time in science as masses of data pour in, promising new insights. But how can we find meaning in these terabytes? To search successfully for new science in large datasets, we must find unexpected patterns and interpret evidence in ways that frame new questions and suggest further explorations. Old habits of representing data can fail to meet these challenges, preventing us from reaching beyond the familiar questions and answers.

To extract new meaning from the sea of data, scientists have begun to embrace the tools of visualization. Yet few appreciate that visual representation is also a form of communication. A rich body of communication expertise holds the potential to greatly improve these tools. We propose that graphic artists, communicators and visualization scientists should be brought into conversation with theorists and experimenters before all the data have been gathered. If we design experiments in ways that offer varied opportunities for representing and communicating data, techniques for extracting new understanding can be made available.

Visual representation is familiar in data-intensive fields. Years before a detector is built for a facility such as the Large Hadron Collider near Geneva, for example, physicists will have pored over simulations. They examine how important events will 'look' in the displays that reveal and communicate what is going on inside the machine. Such discussions tend to take place within the visual conventions of a field. But perhaps conversations might be broadened to consider alternative representations of the same data. These might suggest other approaches to collecting, organizing and querying data that will maximize the transparency of experimental results and thus aid intuition, discovery and communication.

Unfortunately, visualization experts and communicators are often consulted only after data are organized and stored, in the hope that



they will create effective computer displays, slides and figures for publication. Meanwhile, they may be developing their tools in isolation, kept at arm's length by scientists who are busy getting their experiments done. Opportunities for useful dialogue are thus squandered.

When scientists, graphic artists, writers, animators and other designers come together to discuss problems in the visual representation of science, such as at the Image and Meaning workshops run by Harvard University (www.imageandmeaning.org), it becomes clear

those run by the US National Science Foundation's Picturing to Learn project (www.picturingtolearn.org), teach us that attempting to visually communicate scientific data and concepts opens a path to understanding. When science and design students collaborate, their drive to understand one another's ideas pushes them to create new ways of seeing science. Investment in visual communication training for young scientists will pay off handsomely for any data-intensive discipline.

The ingrained habits of highly trained scientists make them rarely as adventurous as these young minds. We think we are on the path to insight when shading reveals contours in 3D renderings, or when bursts of red appear on heat maps, for example. But the algorithms used to produce the graphics may create illusions or embed assumptions. The human visual system creates in the brain an apparent understanding of what a picture represents, not necessarily a picture of the underlying science. Unless we know all the steps from hypothesis to understanding — by conversing with theorists, experimentalists, instrument and software developers, visualization



Discussing visual communication before designing experiments may reveal new science.

that representations repeatedly fail to communicate understanding or address obvious questions about the underlying data. A three-dimensional volume rendering may give no hint of important uncertainties or data gaps; solid surfaces or sharp edges may suggest data where they do not exist. A graphic artist might propose ways to reveal gaps or deviations from expectation early in an experiment, guiding subsequent data collection or highlighting new avenues of enquiry. When we asked Harvard University chemist George Whitesides to change the geometry of a self-assembled monolayer with clearly delineated hydrophobic and hydrophilic areas to create an image for submission to a journal, he found himself redesigning the experiment, and unexpected science emerged.

Student workshops and exercises, such as

scientists, graphic artists and cognitive psychologists — we cannot be sure whether a display is accurate or misleading.

The greatest opportunity and risk lie in that last step in the path: understanding. Whether verbal or visual, any language that is garbled and inconsistent fails to do its job. Let's talk. Let's all talk. ■

Felice Frankel is senior research fellow in the faculty of arts and sciences at Harvard University, Cambridge, Massachusetts 02138, USA. With G. M. Whitesides, she is co-author of *On the Surface of Things: Images of the Extraordinary in Science*. e-mail: felice_frankel@harvard.edu

Rosalind Reid is executive director of the Initiative in Innovative Computing at Harvard University and former Editor of *American Scientist*.

See Editorial, page 1.

the administrator retorted "We don't care if the theories we write about are right, wrong, seriously flawed, downright ignorant or otherwise really, really bad. We only care that the subject is notable." As someone who believes passionately in the value of scholarship, I find this disdain for expert opinion alarming. One hopes that this is an isolated incident, but it does make one wonder whether Google's Knol project or Sanger's Citizendium, which are collecting encyclopaedic articles contributed by named experts, may eventually generate sufficient critical mass to compete with Wikipedia.

Zittrain's book contains more of interest, notably his discussion of 'Privacy 2.0', which includes Scott McNealy's famous quote: "You have zero privacy anyway. Get over it." However, I fear that his concluding example of the One Laptop Per Child project as embodying "both the promise and challenge of generativity" could prove unfortunate if it fails to live up to expectations. He quotes Nicholas Negroponte, the project's founder, as saying "The hundred-dollar laptop is an education project. It's not a laptop project." Negroponte is quoted as saying exactly the opposite in the recent 'resignation' blog of Ivan Krstic, one of the project's early supporters.

Shirky's enjoyable book *Here Comes Everybody* has insights beyond examples of the power of the web and social networking tools. The collapse of transaction costs for people to join or create groups, he claims, is the driving force behind the Internet revolution. The book describes how mass amateurization has displaced media professionalism, and emphasizes that the web is not merely a new competitor but a whole new ecosystem. Publishers still control the production of print articles and books, but this is increasingly irrelevant now that the costs of print reproduction and distribution have disappeared as a result of the web. On websites such as iStockphoto, photographs by amateurs can be found and purchased as easily as those of professionals, removing any distinction.

Shirky explains how "the difference between communication tools and broadcasting tools was arbitrary, but the difference between conversing and broadcasting is real." Many web postings are tedious because they are meant for a few friends, but the web allows thousands of others to listen in. Shirky discusses how the web is creating new models of organization, yet the arguments are not new — I was reminded of Ricardo Semler's paper on 'Managing Without Managers', published in the *Harvard Business Review* back in 1989.

Shirky provides the most detailed discussion of the open-source software movement of the three books, although none undertakes a thorough analysis. Shirky states that "Open-source software has been one of the successes of

the digital age" and characterizes one of its key features as allowing "failure for free". Software companies need to be conservative and to minimize risk; by contrast, the open-source community can explore a vast landscape of different ideas. Leadbeater believes the community behind development of the Linux operating system is "the most impressive example of sustained We-Think", although he acknowledges the paradox of companies such as Google, IBM and HP making money using open-source software. All three authors subscribe to

the idea that open-source software is produced by an unpaid army of volunteers from around the world. The situation is not so simple.

Open-source projects can be divided into two clusters, 'money-driven' and 'community-driven', according to a 2006 paper by Marco Iansiti and Gregory L. Richards of Harvard Business School. The first type has received billions of dollars in investment from vendors over the past decade; for example, more than 70% of the Linux kernel development is carried out by professional software developers. Community-driven open-source projects constitute well over 95% of the 150,000 or so projects in the SourceForge open-source software repository, but the vast majority of these have a handful of users and developers. Nonetheless, Shirky points out, the tendency of open-source projects to fail is also the movement's strength: "Open source is a profound threat, not because the open-source ecosystem is outsucceeding commercial software but because it is outfailing them."

"I would like to think that the web will change the world, but it seems naive to think that it will change human nature."

Leadbeater's *We-Think* is the least convincing book. He gives interesting examples of social networks and a fascinating survey of the origins of 'We-Think', attributing the idea to pioneers such as Doug Engelbart, inventor of the computer mouse, the electronics enthusiasts of Silicon Valley's Homebrew Computer Club and radical philosophers such as Ivan Illich. Leadbeater

claims that the web has the potential to "spread democracy, promote freedom, alleviate inequality and allow us to be creative together" and claims that "community and conver-

sation are the roots of creativity". Yet I find it difficult to take seriously his basic premise that mass participation will generate collaborative creativity. I concede that collaboration between specialists will become more important as we attack challenging global problems. In my experience, creativity and inspiration are the rare gifts of individuals, following much scholarship and hard work. Although I would like to think that the web will change the world for the good — and I am sure that it will in some ways — it seems naive to think that it will cause a fundamental change in human nature.

These three books contain much that is perceptive, informative and downright silly, much like the Internet itself. It is this 'generativity' that they celebrate. ■

Tony Hey is Corporate Vice President for External Research, Microsoft Research, One Microsoft Way, Redmond, Washington 98052, USA.
e-mail: tony.hey@microsoft.com

All opinions in this review are personal and do not represent the views of the Microsoft Corporation.

Virtual similarities

Coming of Age in Second Life: An Anthropologist Explores the Virtually Human
by Tom Boellstorff

Princeton University Press: 2008. 328 pp.
\$29.95, £17.95

In his book *Secondary Worlds*, W. H. Auden wrote that "present in every human being are two desires, a desire to know the truth about the primary world ... and the desire to make new secondary worlds of our own or, if we cannot make them ourselves, to share in the secondary worlds of those who can". Auden, in 1968, was writing about literature, not cyberspace, but his thoughts help explain

why virtual worlds are popular today.

The conflicts that arise from this desire to live in both a primary world and a secondary world removed from physical reality are examined in *Coming of Age in Second Life*. Anthropologist Tom Boellstorff paints an ethnographic portrait of the online virtual world, Second Life, that is fully immersed in its subject. To prove that virtual worlds are cultures in their own right, Boellstorff conducted all his research from within Second Life, using the ethnographer's toolkit of interviews, focus groups and participant observation. Unlike other studies that take an outside perspective, he made no attempt to make real-life contact with his fellow residents.

Some may argue that it is not possible to

understand a person's virtual life without knowing their actual-life history. Online, residents can mask their identity, including their race, gender and age. Some adopt multiple virtual bodies — avatars — and some avatars are controlled by more than one person. Does Boellstorff's approach have any value? The author argues that residents have created a culture in which it is not necessary to know a person's true identity to engage in meaningful interactions. To understand how these relationships work, an anthropologist should not need any additional information, and behaviours can be better examined from the same viewpoint as the subjects of the study.

The gap between the virtual and the physical, and its effect on the ideas of personhood and relationships, is the most interesting aspect of Boellstorff's analysis. For many residents, having a separate embodiment in cyberspace is liberating. There, they are free to be the people they imagine themselves to be, no longer held back by real attributes or attitudes. This liberation was evident in transgender people experimenting with a new lifestyle in Second Life before making decisions in the actual world.

Freedom is the primary attraction for many residents. When Linden Lab, the software company behind Second Life, pondered whether to introduce voice communication to the platform, in addition to the existing textual chat, it provoked widespread riots within the virtual world. The sound of a real voice would have closed the gap between the real person and the virtual avatar to an unacceptable degree.

The real–virtual disparity can bring a painful



T. BOELLSTORFF

Tom Boellstorff's anthropologist avatar carries out his research within the virtual world.

distance between virtual friends or lovers. Boellstorff writes of Susan and George, who enjoyed a close relationship conducted solely in the virtual world for a year and a half. When George failed to log in for a whole month, Susan's devastation was real. He may have died in the real world, or might have simply tired of the relationship, but without knowing his true identity, Susan was powerless to find out. Many commentators question the

existence of meaningful relationships within virtual worlds, but Boellstorff demonstrates that the emotional commitment invested makes them just as real and worthy of study.

A portrait of Second Life would not be complete without documenting the more colourful members of its society, such as the 'furries' who are embodied in animal avatars or the virtual sex workers. Given the popular media coverage of Second Life, it would have been easy to focus on these sensational residents, but the closer study of more mundane characters such as Susan and George provides valuable insights. Boellstorff shows that although Second Life culture has its own unique nuances, for the majority of residents it is no more surprising than societies based in the physical world.

During the period of Boellstorff's study, June 2004 to January 2007, the population of Second Life grew from a few thousand to several million, with important software upgrades along the way. Technology moves quickly, and the society portrayed in *Coming of Age in Second Life* may change in the future. Boellstorff's portrayal of a virtual culture at the advent of its acceptance into mainstream life gives it lasting importance, and his methods will be a touchstone for research in the emerging field of virtual anthropology.

David Robson is a writer based in London, UK.
e-mail: d_a_robson@hotmail.com



T. BOELLSTORFF

Focus groups revealed that virtual relationships were just as important as physical ones.

Q&A: Museum's metamorphosis is nearly complete

On the unveiling of the second phase of the Darwin Centre at London's Natural History Museum, **Anna Maria Indrio**, partner at the Scandinavian architectural firm C. F. Møller, explains how the new £78 million (US\$145 million) wing will reveal 20 million of the museum's insect and plant specimens to the public when it opens in September 2009.

What were the design challenges?

The biggest issue was the huge size of the collections, which are among the world's most extensive and treasured. Protecting such a valuable array of 17 million insects and 3 million plants in 3 kilometres of cabinets, showcasing them to the public and ensuring that the design represents the scientists' work was very daunting.

How are the specimens shielded?

A cocoon — representing preservation, protection and nature — forms the inner envelope of the building. Shaped according to mathematical equations, it is 8 storeys high, 65 metres long and is the largest sprayed-concrete curved structure in Europe. The hand-finished surface of ivory-coloured polished plaster resembles a silk cocoon; a series of expansion joints wrap around like silk threads. To emphasize the massive scale of the collections, the giant cocoon can never be seen in its entirety.

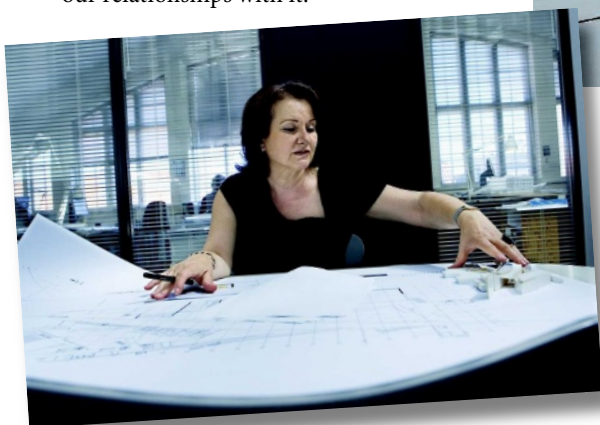
How will the public experience the Darwin Centre?

Around 2,500 visitors per day will journey through a series of exhibits within the cocoon. They will be able to watch scientists at work in glass-fronted laboratories through windows at the end of the cocoon. And through natural-history films, new media and face-to-face encounters with museum scientists, visitors will be inspired to be naturalists — observing the natural world and debating our relationships with it.



The Darwin Centre cocoon designed by Anna Maria Indrio (below left) is the largest sprayed-concrete curved structure in Europe.

C. F. MØLLER; NAT. HIST. MUSEUM



What research facilities are included?

More than 200 scientists will be able to work in the centre at any one time in purpose-built laboratories, doubling the research space at the museum. Its open-plan layout is designed to help the exchange of ideas. The eighth-floor common room will be shared across all the life-science departments.

How will you feel when it opens?

We discovered an astonishingly complex world at the museum, with many layers to the design brief. We will be hugely proud when the scientists start moving in early next year.

Interview by **Joanne Baker**, *Nature's Books & Arts* Editor.

See <http://tinyurl.com/2rftbf> for further information.

In Retrospect: Leibniz's *Protogaea*

The first English translation of Gottfried Leibniz's earth science treatise records the difficulties of understanding our planet before geologists appreciated deep time, **Richard Fortey** discovers.

Protogaea

by Gottfried Wilhelm Leibniz

Translated by Claudine Cohen and
Andre Wakefield

University of Chicago Press: 2008.
204 pp. \$55.

It is something of a game among historians to try and detect the earliest hints of a major scientific breakthrough in a little-known work discovered through recondite scholarship. Charles Darwin's supposed debt to his grandfather Erasmus is an example, or maybe geologist Charles Lyell's insufficient acknowledgement of the early geological work of Nicolaus Steno. When the savant in question is Gottfried Wilhelm Leibniz (1646–1716) — the man who developed calculus independently of Isaac Newton — hidden insights might genuinely be anticipated. Here was a prolific thinker of range and profundity. His *Protogaea*, a posthumously published 1749 treatise on earth sciences, has now been translated from its original Latin into English for the first time, and bears a title that chimes with our current concerns about global ecosystems. What did the great man make of the history of Earth?

As he states in the book, Leibniz intended to develop “the seeds of a new science called natural geography”. The original text would have been readily comprehensible to his contemporaries, and must surely have seeped into subsequent thoughts about geology. Indeed, had his book been a more complete account, ‘geology’ might have been a stillborn term.

As it is, Leibniz picks out facts derived from his own observations, from his network of correspondence with other natural philosophers, and from his wide reading of those he regards as trustworthy observers. His text briefly touches on many geological phenomena, from the formation of mountains to the origin of minerals and particularly fossils, which in this new translation are well illustrated by reproductions of the original contemporary woodcuts. Leibniz was unusually scathing for his time about those who ‘see’ miraculous religious resemblances in natural objects, and wrote: “credulity fills in the rough outlines shaped by accident”. Richard Dawkins could not have put it better.

Presciently, Leibniz is equally clear about the organic nature of many fossils: fish preserved in slates are exactly that and not mere ‘games



The ‘unicorn’ of Quedlinburg fooled even Leibniz.

of nature’. He was reacting against philosophers such as Athanasius Kircher, who “claim the great architect, as if in jest, had imitated the teeth and bones of animals, shells or snakes”. Leibniz was certain that God had much more serious purposes than planting simulacra in the rocks. *Glossopetrae*, or ‘tongue stones’, are shark’s teeth, he states, nothing more or less. He recognizes ammonites and other fossil shells as having more than a passing similarity to their living relatives. The philosophical mind at work in these passages is of a modern, sceptical cast, highlighting that Leibniz was well ahead of most of his contemporaries.

Even Leibniz is occasionally credulous. One illustration in the book shows the unicorn of Quedlinburg, a chimera of several mammals that was ‘discovered’ in 1663. “The horn, together with the head, several ribs, dorsal vertebrae and bones were brought to the town’s serene abbess”, Leibniz confides, evidently deeming the words of this particular local eyewitness reliable. Nonetheless, he takes on board the field examples described by Steno that show how a sequence of strata revealed something of Earth’s history. Scientific narrative was only a step away.

When considering the origin of minerals, Leibniz has an intuitive sense that a kind of natural cookery is involved: “One is thus inclined to suspect that nature, using volcanoes as furnaces and mountains as alembics, has accomplished in her mighty works what we play at with our little examples [in laboratories].” That the furnaces of the ‘chymist’ might simulate Earth’s processes is a hope that still drives research into petrology and geochemistry today.

Why then did Leibniz’s shrewd observations fail to move geology significantly towards becoming a mature science? For all its insights, *Protogaea* does not seem to a modern geologist like the natural ancestor of Lyell’s *Principles of Geology*. The missing ingredient is an awareness of geological time. Leibniz did not place the biblical timescale centrally in his science — he was actually very restrained in invoking the Creator. A short timescale was simply a given, so widely accepted that he did not have to restate it. Even Leibniz’s evident awareness of events such as major incursions of the sea over what is now dry land did not challenge his view. Geology without time is rather like chemistry without elements: a collection of plausible narratives is possible; a rational basis for predictive science is not.

More mundanely, it was also difficult to travel in the eighteenth century. Leibniz had to rely on the observations of others simply because wide-scale fieldwork was almost impossible. Armchair speculation was inevitable, despite Leibniz’s careful affirmation of his own observations. The true complexity of Earth’s history did not begin to be exposed until French geologists started work in the Auvergne and the Paris basin, and until James Hutton developed his theory of deep geological time in Scotland. The improvement of roads and canals, and then the advent of railways, allowed for different local geological narratives to be stitched together. Earth science has subsequently developed to recognize our planet as an interconnected system that has evolved over billions of years. Leibniz’s world is an incomplete patchwork of local stories. ■

Richard Fortey is a research associate at the Natural History Museum, Cromwell Road, London SW7 5BD, UK. He is author of *The Earth: An Intimate History*.
e-mail: r.fortey@nhm.ac.uk

ESSAY

The Harvard computers

The first mass data crunchers were people, not machines. **Sue Nelson** looks at the discoveries and legacy of the remarkable women of Harvard's Observatory.

A photograph taken at the Harvard Observatory in Cambridge, Massachusetts, circa 1890, features eight women in what looks like a Victorian-



style sitting room. They wear long skirts, have upswept hair and are surrounded by flowered wallpaper and mahogany tables. At first glance they seem to be sampler stitching or reading. In fact these 'human computers' are analysing photographs of the heavens, cataloguing stars.

When cameras were first attached to telescopes, with the ability to capture the image of thousands of stars on a single photographic plate, people were needed to trawl through these new data. Observatories hired 'computers' — a term used for human processors since the early 1700s — to do the painstakingly repetitive work of measuring the brightness, position and colours of these stars.

From the 1880s until the 1940s, the Harvard College Observatory amassed half a million photographic glass plates, weighing around 300 tonnes and holding images of tens of millions of stars. A team of women trawled through these photos with nothing more than magnifying glasses — often for little pay and with no scientific training.

Despite these unpromising conditions, the 'Harvard computers', who worked from the end of the nineteenth century to the mid-1920s, made tremendous contributions to astronomy. They determined how to calculate the vast distances from Earth to the stars, and developed star classification systems that are still used today. From photos taken of the northern and southern skies, from observatories in Cambridge, New Zealand and Peru, they produced an astronomical gold mine of data.

These women were proof that 'people power', even from those with no formal training, is capable of great things. It is a trend that continues today: volunteers are recruited from the general population and taught to spot objects of interest to astronomers, from the tracks of interstellar dust left in a spacecraft's collector, to the direction of spin of a spiral galaxy. With Harvard now working to digitize its photographic plates, the same pictures of stars scrutinized by the Harvard computers may soon be available to many more, equally curious, eyes.

Working with the repetitive and often indistinct photographs collected at places such as the Harvard Observatory required patience,



HARVARD COLLEGE OBSERVATORY

Williamina Fleming stands in the centre of the Harvard computers as Edward Pickering looks on.

attention to detail and stamina. Most of the plates are negatives; stars appear as fine grey or black spots against a clear background. There are also several thousand spectral plates, in which starlight has been split by a prism before being captured. These look like nothing more than smudged pencil marks a few millimetres wide; under a magnifying glass the smudge turns into a barcode, revealing information about the chemical composition and temperature of the stars.

Patience personified

In 1901, William Elkin, the director of Yale Observatory, expressed a view typical of the time as to who was best suited for this work. "I am thoroughly in favour of employing women as measurers and computers," he said. "Not only are women available at smaller salaries than are men, but for routine work they have important advantages. Men are more likely to grow impatient after the novelty of the work has worn off and would be harder to retain for that reason."

Edward Pickering, the Harvard College Observatory director in 1877–1919, famously

said that the computing work at his observatory was so easy that even his "Scotch maid" could do it. This was Williamina Fleming, a schoolteacher from Dundee who had emigrated to America with her husband in 1878. A year later, abandoned and pregnant, she secured a job as Pickering's maid and housekeeper. She was soon working for him at the observatory part-time as one of his first computers.

Pickering's apparently disparaging remark about his maid belies the fact that he spotted and nourished the untapped potential in many intelligent women who worked for him. Fleming was obviously bright and Pickering recognized this; by 1881, at the age of 24, she was appointed a full-time staff member of Harvard. Seven years later, she assumed responsibility for the increasing number of photographic plates, editing publications from the observatory and hiring new computers. During her time at Harvard, Fleming examined thousands of spectra and catalogued more than 10,000 stars.

Fleming helped Pickering to devise his hydrogen-based stellar classification system, which ranked stars according to the strength of a hydrogen spectral line — A for the

strongest, then B and so on. She also played a crucial part in the discovery of the spectral peculiarities of white dwarfs. Fleming was appointed Harvard's Curator of Astronomical Photographs in 1899 — the first woman so appointed — and eight years later she became the first female American citizen elected to the British Royal Astronomical Society. During her career it is estimated that she examined around 200,000 photographic plates.

Pickering's 'harem', as they have been called, could have earned more per hour doing menial work in the local mill town. But the observatory appealed to intelligent and educated women, including researchers and graduates from the new all-female colleges, who were keen to find patterns in the data and draw interesting conclusions from them. Pickering allowed the women to do their own research in their spare time and their names were often cited as co-authors in scientific papers. He encouraged them to give talks, and to be recognized as astronomers in their own right.

Annie Jump Cannon, for example, measured and recorded the colours of 300,000 stars, classifying them into spectral groups at a rate of up to 300 an hour. Her true achievement came when she was tasked with finding a more meaningful way of arranging the star categories in Pickering's hydrogen-based system. Cannon developed a reordering and simplification that ranked stars in order from the bluest and hottest to the coolest, red ones. The sequence — O, B, A, F, G, K, M (remembered by the mnemonic Oh Be A Fine Girl, Kiss Me) — remains in use today.

Henrietta Swan Leavitt, a college graduate, joined Pickering as a research assistant. She studied Cepheid variable stars — whose light brightens and dims over periods ranging from hours to years — and found that these periods were related to the bodies' intrinsic brightness in a predictable way. This relationship, once calibrated, allowed astronomers to determine such a star's distance, based on its true brightness as compared to its apparent brightness as seen from Earth. This 'Cepheid distance scale' became Edwin Hubble's 'yardstick to the Universe', eventually allowing him to discover that the Universe was expanding. It is still crucially important in distance measurements today.

Leavitt died in 1921 of cancer, four years before a letter was sent notifying her that she had been nominated for a Nobel prize — an honour that had to be withdrawn because it could not be awarded posthumously. In his 2005 book *The Discoveries*, physicist Alan

Lightman nominates the Cepheid distance scale as one of the most important breakthroughs in twentieth-century science. To celebrate the centennial of Leavitt's pioneering work, the Harvard College Observatory is holding a symposium in her honour this November.

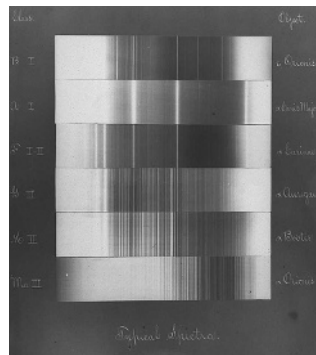
More advanced cameras and digital photography removed the need for most photographic plates in the 1980s. But the sheer number of plates collected at Harvard means that even now there are stars that have yet to be analysed. Because of this untapped potential, and the historical significance of the computers' work, the DASCH (Digital Access to a Sky Century at Harvard) project is hoping to scan all the half-million plates over the next four years; the team has so far scanned more than 3,000 plates and is trying to raise US\$4 million to complete the work.

Pattern recognition

The primary motivation of this work is not historical preservation, but to dig deeper into the data. Requests have already been filed by researchers for access to the DASCH's digitized collection. One Harvard graduate student, Sumin Tang, working with DASCH principal investigator Josh Grindlay, recently found something unusual after studying just 500 of the digitized plates: a star that brightened by nearly a factor of two over 20 years and then levelled off, retaining its brightness for 60 years. This new type of variable star will be observed further using a telescope in Arizona in the years to come.

The sheer magnitude of data expected for the entire DASCH project (around 1,500 terabytes) lends itself well to analysis by today's high-speed computers. Yet there remain some areas in which the human eye is better than a modern computer, particularly in the realm of pattern recognition. Classifying stars or galaxies, like classifying species, is still something most easily done by people with knowledge and a knack.

In some cases, the best results come from the use of many eyes — even if they are untrained. NASA's 2004 Stardust mission, for example, gathered dust particle samples on an aerogel



Spectra (top) pulled from star plates (above), are only millimetres wide.

collector during a rendezvous with a comet. One of the greatest challenges then lay in finding the tracks of tiny interstellar particles — rare pieces of dust that came from distant stars — among the more common tracks left by comet particles. These tracks, at a millionth of a metre across, were a devil to find on the 1,000-square-centimetre collector.

In 2006, the Planetary Society and the University of California, Berkeley, launched the Stardust@Home project. Just as the Harvard computers were trained to analyse stellar spectra, members of the public were trained, via online tutorials, to scan photos of the gel on their computer screens and identify possible tracks. No computer program exists that can

do this as well as the human eye.

In the first phase of the project, 23,000 volunteer 'dusters' searched nearly 40 million images, flagging any photos of possible interest for trained scientists. Without public help, it would have taken the team at least 20 years to locate tracks. With their help, it took months. Phase two is currently under way, with dusters searching for particles in photos of a higher magnification.

Inspired by the Stardust project, Galaxy Zoo went online in 2007. This project used volunteers to classify spiral and elliptical galaxies from images taken by the Sloan Digital Sky Survey. Within six months of operation, volunteers had identified more than 500 overlapping galaxies — when a galaxy appears behind or in front of another from the point of view of an observer on Earth — of which only 20 were previously known to astronomers. Recently, a unique object containing the hot gas of a normal galaxy but without any stars was discovered through the project by a Dutch primary school teacher; a paper describing the find is in the works.

Modern astronomy could not be done without supercomputers crunching epic quantities of data. It is nonetheless worth remembering the value of the human mind in spotting complex patterns or following a hunch — as proved by the persistent, repetitive and inspiring work of the pioneering women at Harvard. ■

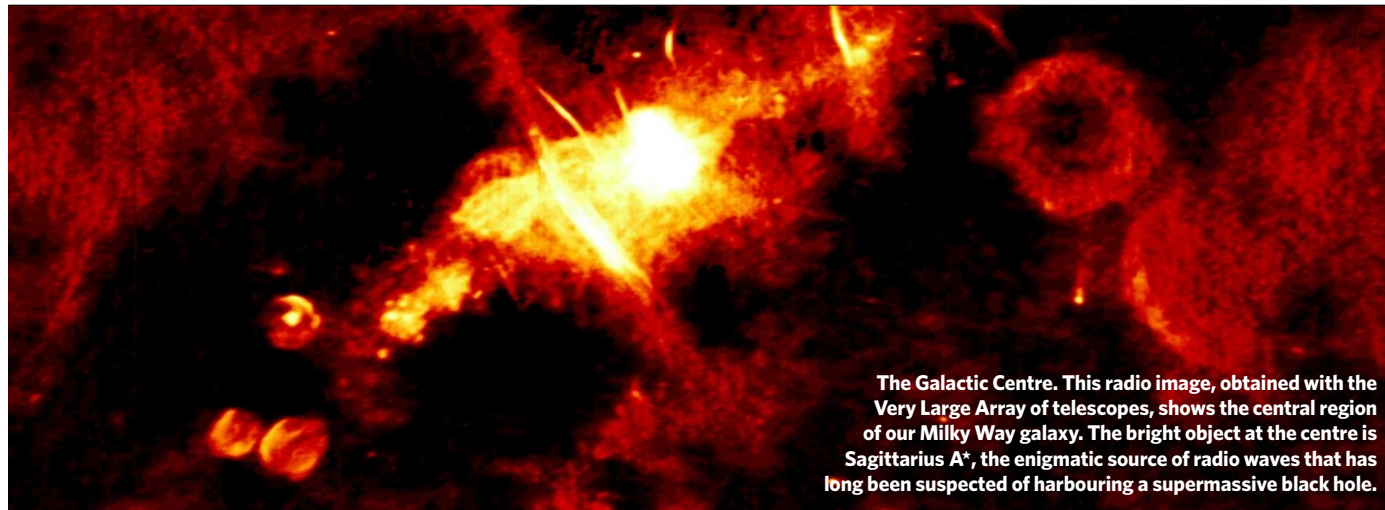
Sue Nelson is a writer and broadcaster living in Hertfordshire, UK.

e-mail: sue.nelson@zen.co.uk

See Editorial, page 1.

"There remain some areas in which the human eye is better than a modern computer."

NEWS & VIEWS



KASSIM, LAROSA, LAZIO & HYMAN / NAVAL RESEARCH LAB.

The Galactic Centre. This radio image, obtained with the Very Large Array of telescopes, shows the central region of our Milky Way galaxy. The bright object at the centre is Sagittarius A*, the enigmatic source of radio waves that has long been suspected of harbouring a supermassive black hole.

ASTROPHYSICS

Bringing black holes into focus

Christopher S. Reynolds

Do black holes exist? Observations at the finest resolution so far indicate that only gross deviations in the behaviour of gravity from that predicted by general relativity can invalidate the case that they do.

It is believed that the centre of essentially every galaxy, including our own, plays host to a supermassive black hole. In a small fraction of galaxies, large quantities of gas rain down into these giant black holes, causing the black hole to grow while releasing enough energy within the central few light hours of the galaxy to outshine all of the galaxy's stars thousands of times over. This is more than a mere cosmic firework show; the energy released as the black hole grows can shape or even shut off the processes by which the galaxy itself forms. In other words, supermassive black holes may well be the safety valve that regulates galaxy formation, preventing galaxies from growing too big too fast. But although they are rapidly becoming a standard part of our model of how galaxies form and evolve, it is important to step back and ask just how strong is the case that these monster black holes actually exist.

On page 78 of this issue, Doeleman *et al.*¹ report new observations of Sagittarius A* (Sgr A*), the enigmatic source of radio waves at the centre of our Galaxy² that has long been suspected as signposting our very own supermassive black hole. These new data have allowed the authors, for the first time, to detect structure in the radio emissions from Sgr A* on scales as small as 50 million kilometres. The diameter of our Galaxy's black hole (which has a mass 4 million times that of the Sun) is expected to be approximately 12 million to

24 million kilometres. But the strong bending of light rays within the gravitational field of the black hole will double the apparent size of the event horizon, the boundary of the area around the black hole from which nothing, not even light, can escape. Thus Doeleman and colleagues' observations have finally brought us to the threshold of imaging horizon-scale structures — a holy grail of radio astronomy.

With the new data, the authors have attained a resolution of about 40 microarcseconds (about one-hundred millionth of a degree), five times better than the best previous measurement³. This advance has been made possible by extending the technique of very long baseline interferometry (VLBI) to shorter radio wavelengths — indeed, into the microwave region of the electromagnetic spectrum. In VLBI, data from radio telescopes spread across the globe are combined to produce vastly superior image resolution than can be achieved by any one telescope; but this process requires keeping track of the precise phase of the incoming waves. This technological feat becomes increasingly challenging as the wavelength of the waves is decreased in the search for superior resolving power. The observation reported by Doeleman *et al.*¹, made with telescopes in Arizona, California and Hawaii, is one of the first to exploit VLBI with 1.3-mm waves.

Black holes are truly bizarre objects. Einstein's theory of general relativity tells us that

they are objects in which gravity has run amok, cutting off a region of space (inside the event horizon) from the outside Universe. Inside the event horizon, theory predicts the existence of regions in which densities and temperatures climb to such extreme values that all currently understood laws of physics break down. These new results¹ put us a step closer to confirming that nature really is this anarchistic. Assuming that the central object must be smaller than the surrounding 'cloud' of radio-emitting gas that we see, the case for a black hole looks compelling. Even a 4-million-solar-mass boson star, an exotic hypothetical object sometimes discussed as an alternative to black holes⁴, will be much larger in extent than the 50-million-kilometre limit implied by Doeleman and colleagues' data. Given these data, only gross deviations in the behaviour of gravity itself from the behaviour predicted by general relativity can invalidate the case for black holes.

Efforts to improve the sensitivity and imaging ability of millimetre-wavelength VLBI promise further dramatic advances in our understanding of Sgr A*. For example, future studies will reveal effects related to the spin of the black hole. Although still the subject of intense research, the complex gas flows close to a black hole can be strongly affected by the tornado-like motion of space-time close to a spinning black hole⁵, as can the appearance of the 'shadow' of the event horizon⁶. Characterizing these phenomena will

allow us to determine the spin rate of the black hole, offering a window into its long cosmic history. Did it grow through the successive mergers of smaller black holes as galaxies came crashing together? Or did it grow through the accretion of gas and, if so, did it snack on gas hundreds of times⁷ or feast just once or twice? The spin of the black hole encodes, albeit crudely, this history and may be one of our best handles for understanding the evolution of this, and other, supermassive black holes⁸.

We have entered a new era, one in which we can now directly image structure at the event horizon of a black hole. As the VLBI array capable of millimetre resolution is expanded and its sensitivity increased, the distorted

world at the edge of the black hole will literally come into focus.

Christopher S. Reynolds is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA.

e-mail: chris@astro.umd.edu

1. Doeleman, S. S. *et al.* *Nature* **455**, 78–80 (2008).
2. Balick, B. & Brown, R. L. *Astrophys. J.* **194**, 265–270 (1974).
3. Shen, Z. Q., Lo, K. Y., Liang, M.-C., Ho, P. T. P. & Zhao, J.-H. *Nature* **438**, 62–64 (2005).
4. Torres, D. F., Capozziello, S. & Lambiase, G. *Phys. Rev. D* **62**, 104012 (2000).
5. Krolik, J. H., Hawley, J. F. & Hirose, S. *Astrophys. J.* **622**, 1008–1023 (2005).
6. Reynolds, C. S. *Nature* **438**, 32–33 (2005).
7. King, A. R. & Pringle, J. E. *Mon. Not. R. Astron. Soc.* **377**, L25–L28 (2007).
8. Volonteri, M. *et al.* *Astrophys. J.* **620**, 69–77 (2005).

IMMUNOLOGY

Oxysterols hold T cells in check

Christopher K. Glass and Kaoru Saijo

The oxysterol-dependent gene transcription factor LXR β restricts premature expansion of T cells by limiting cellular cholesterol levels. This pathway might be a pharmacological target for regulating immune responses.

Adaptive immune responses depend on the activation and expansion of specific subsets of white blood cells such as T cells in response to antigens. Although it has long been appreciated^{1,2} that such proliferative responses are linked to the uptake and *de novo* synthesis of cholesterol for membrane formation, whether cholesterol-efflux pathways can limit cell division has not been considered. Reporting in *Cell*, Bensinger *et al.*³ provide evidence that cholesterol efflux is indeed used to inhibit the proliferation of resting T cells — T cells that have not yet encountered an antigen.

Intracellular cholesterol levels are tightly regulated by two complementary pathways that are mediated by the gene transcription factors SREBP and LXR. The SREBP-dependent pathway induces the expression of proteins that are required for cholesterol biosynthesis and uptake, such as HMG CoA reductase and the LDL receptor, thereby increasing cellular cholesterol levels⁴. This pathway is regulated by feedback inhibition, but it cannot eliminate excess cholesterol once this lipid has accumulated. That task is in part accomplished by cholesterol-efflux pathways that are regulated by LXR α and LXR β — members of the nuclear-receptor superfamily. Oxidized cholesterol derivatives (oxysterols) activate LXRs⁵, which then control the transcription of genes that have diverse roles in cholesterol homeostasis and innate immunity⁶. For example, in many cell types LXRs promote a reduction in cellular cholesterol levels by inducing the expression of the ABCA1 and ABCG1 transporters, which mediate cholesterol efflux from the cell to extracellular acceptors.

Bensinger *et al.*³ find that mice lacking both LXR α and LXR β develop enlarged spleens and lymph nodes owing to increased numbers of immune B cells and T cells. T cells express only LXR β , and its loss seems to be responsible for the increased numbers of these cells.

The authors find that LXR β -deficient T cells exhibit higher rates of proliferation in response to antigens, both *in vitro* and *in vivo*, and that in mutant mice lacking mature T cells in their lymphoid organs, these cells more efficiently repopulate the lymphoid organs than do normal T cells.

Bensinger *et al.* find that cholesterol levels in dividing T cells are determined by the reciprocal regulation of the LXR and SREBP transcriptional programs. They show that activation of T cells by antigens results in increased activity of the SREBP pathway and simultaneously reduced LXR activity. These changes alter the balance of cholesterol homeostasis to allow membrane formation and cellular proliferation (Fig. 1).

Intriguingly, inhibition of LXR signaling seems to be — at least in part — due to the induction of the oxysterol-metabolizing enzyme SULT2B1, which inactivates these natural LXR ligands by adding sulphate groups to them⁷. Synthetic LXR ligands also potentially inhibit the proliferative responses of normal T cells to antigens, overriding the effect of eliminating natural LXR ligands³. This effect is lost in T cells lacking the ABCG1 transporter, indicating a direct link between the antiproliferative activities of LXR ligands and cholesterol efflux. So in the absence of antigens, LXR-mediated limitation of cholesterol availability seems to serve as a checkpoint to restrict T-cell proliferation.

Several questions emerge from this study³. Of immediate interest is determining the extent to which these observations apply to

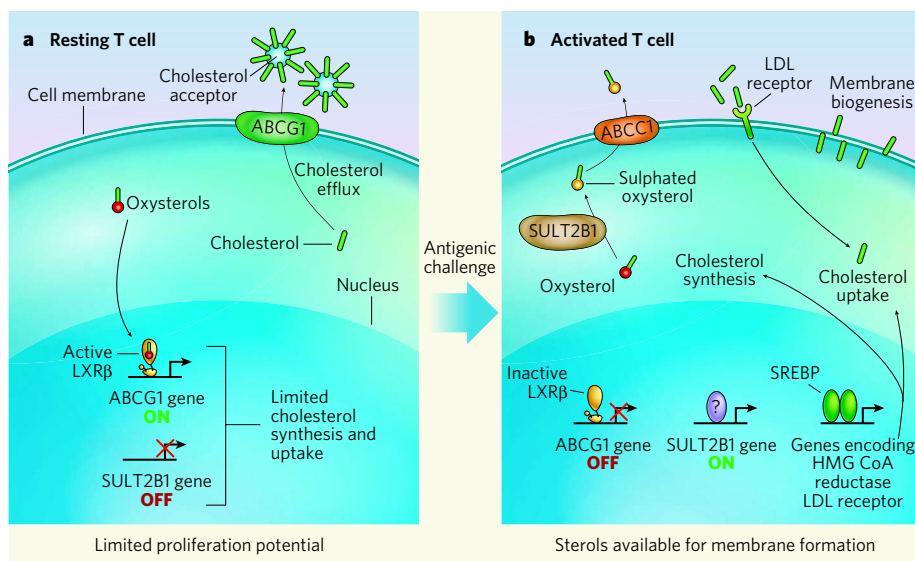


Figure 1 | Response of T cells to a shift in cholesterol balance. **a**, Bensinger *et al.*³ find that, in resting T cells, oxysterols activate the transcription factor LXR β , which in turn leads to increased expression of the ABCG1 transporter for cholesterol transfer out of the cell. In these cells, the activity of the SREBP pathway, which favours *de novo* cholesterol synthesis and cholesterol uptake, is also low. Consequently, there is a shortage of sterols for membrane biogenesis, which is required for effective proliferative responses. **b**, Antigenic challenge activates T cells, resulting in upregulation of the SREBP pathway. By increasing the expression of genes encoding HMG CoA reductase and the LDL receptor, this enhances cholesterol uptake and synthesis. Moreover, increased activity of the oxysterol-metabolizing enzyme SULT2B1 and the ABCG1 transporter eliminates oxysterol ligands of LXR β , leading to reduced activity of cholesterol-efflux pathways. Internalized and newly synthesized cholesterol can now be used for membrane biogenesis, leading to cell proliferation.

allow us to determine the spin rate of the black hole, offering a window into its long cosmic history. Did it grow through the successive mergers of smaller black holes as galaxies came crashing together? Or did it grow through the accretion of gas and, if so, did it snack on gas hundreds of times⁷ or feast just once or twice? The spin of the black hole encodes, albeit crudely, this history and may be one of our best handles for understanding the evolution of this, and other, supermassive black holes⁸.

We have entered a new era, one in which we can now directly image structure at the event horizon of a black hole. As the VLBI array capable of millimetre resolution is expanded and its sensitivity increased, the distorted

world at the edge of the black hole will literally come into focus.

Christopher S. Reynolds is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA.

e-mail: chris@astro.umd.edu

1. Doeleman, S. S. *et al.* *Nature* **455**, 78–80 (2008).
2. Balick, B. & Brown, R. L. *Astrophys. J.* **194**, 265–270 (1974).
3. Shen, Z. Q., Lo, K. Y., Liang, M.-C., Ho, P. T. P. & Zhao, J.-H. *Nature* **438**, 62–64 (2005).
4. Torres, D. F., Capozziello, S. & Lambiase, G. *Phys. Rev. D* **62**, 104012 (2000).
5. Krolik, J. H., Hawley, J. F. & Hirose, S. *Astrophys. J.* **622**, 1008–1023 (2005).
6. Reynolds, C. S. *Nature* **438**, 32–33 (2005).
7. King, A. R. & Pringle, J. E. *Mon. Not. R. Astron. Soc.* **377**, L25–L28 (2007).
8. Volonteri, M. *et al.* *Astrophys. J.* **620**, 69–77 (2005).

IMMUNOLOGY

Oxysterols hold T cells in check

Christopher K. Glass and Kaoru Saijo

The oxysterol-dependent gene transcription factor LXR β restricts premature expansion of T cells by limiting cellular cholesterol levels. This pathway might be a pharmacological target for regulating immune responses.

Adaptive immune responses depend on the activation and expansion of specific subsets of white blood cells such as T cells in response to antigens. Although it has long been appreciated^{1,2} that such proliferative responses are linked to the uptake and *de novo* synthesis of cholesterol for membrane formation, whether cholesterol-efflux pathways can limit cell division has not been considered. Reporting in *Cell*, Bensinger *et al.*³ provide evidence that cholesterol efflux is indeed used to inhibit the proliferation of resting T cells — T cells that have not yet encountered an antigen.

Intracellular cholesterol levels are tightly regulated by two complementary pathways that are mediated by the gene transcription factors SREBP and LXR. The SREBP-dependent pathway induces the expression of proteins that are required for cholesterol biosynthesis and uptake, such as HMG CoA reductase and the LDL receptor, thereby increasing cellular cholesterol levels⁴. This pathway is regulated by feedback inhibition, but it cannot eliminate excess cholesterol once this lipid has accumulated. That task is in part accomplished by cholesterol-efflux pathways that are regulated by LXR α and LXR β — members of the nuclear-receptor superfamily. Oxidized cholesterol derivatives (oxysterols) activate LXRs⁵, which then control the transcription of genes that have diverse roles in cholesterol homeostasis and innate immunity⁶. For example, in many cell types LXRs promote a reduction in cellular cholesterol levels by inducing the expression of the ABCA1 and ABCG1 transporters, which mediate cholesterol efflux from the cell to extracellular acceptors.

Bensinger *et al.*³ find that mice lacking both LXR α and LXR β develop enlarged spleens and lymph nodes owing to increased numbers of immune B cells and T cells. T cells express only LXR β , and its loss seems to be responsible for the increased numbers of these cells.

The authors find that LXR β -deficient T cells exhibit higher rates of proliferation in response to antigens, both *in vitro* and *in vivo*, and that in mutant mice lacking mature T cells in their lymphoid organs, these cells more efficiently repopulate the lymphoid organs than do normal T cells.

Bensinger *et al.* find that cholesterol levels in dividing T cells are determined by the reciprocal regulation of the LXR and SREBP transcriptional programs. They show that activation of T cells by antigens results in increased activity of the SREBP pathway and simultaneously reduced LXR activity. These changes alter the balance of cholesterol homeostasis to allow membrane formation and cellular proliferation (Fig. 1).

Intriguingly, inhibition of LXR signaling seems to be — at least in part — due to the induction of the oxysterol-metabolizing enzyme SULT2B1, which inactivates these natural LXR ligands by adding sulphate groups to them⁷. Synthetic LXR ligands also potentially inhibit the proliferative responses of normal T cells to antigens, overriding the effect of eliminating natural LXR ligands³. This effect is lost in T cells lacking the ABCG1 transporter, indicating a direct link between the antiproliferative activities of LXR ligands and cholesterol efflux. So in the absence of antigens, LXR-mediated limitation of cholesterol availability seems to serve as a checkpoint to restrict T-cell proliferation.

Several questions emerge from this study³. Of immediate interest is determining the extent to which these observations apply to

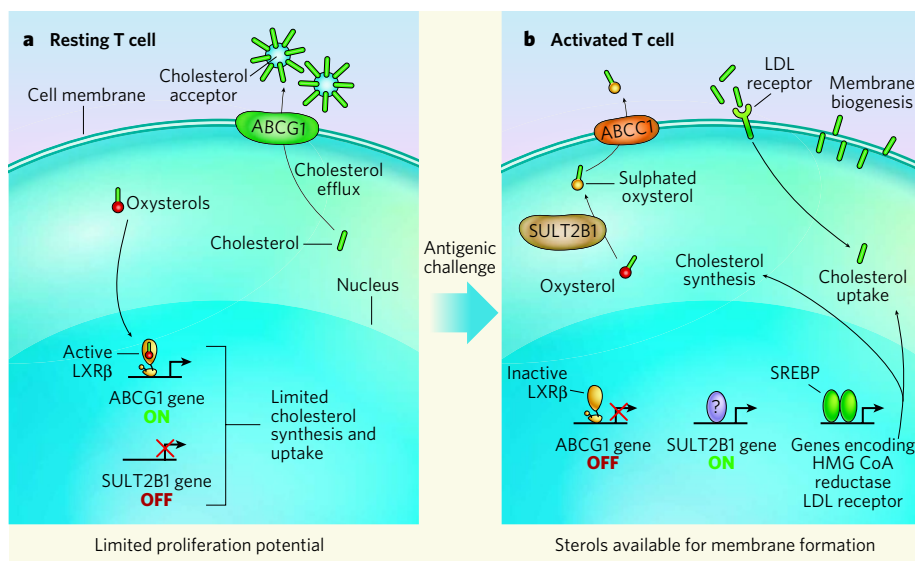


Figure 1 | Response of T cells to a shift in cholesterol balance. **a**, Bensinger *et al.*³ find that, in resting T cells, oxysterols activate the transcription factor LXR β , which in turn leads to increased expression of the ABCG1 transporter for cholesterol transfer out of the cell. In these cells, the activity of the SREBP pathway, which favours *de novo* cholesterol synthesis and cholesterol uptake, is also low. Consequently, there is a shortage of sterols for membrane biogenesis, which is required for effective proliferative responses. **b**, Antigenic challenge activates T cells, resulting in upregulation of the SREBP pathway. By increasing the expression of genes encoding HMG CoA reductase and the LDL receptor, this enhances cholesterol uptake and synthesis. Moreover, increased activity of the oxysterol-metabolizing enzyme SULT2B1 and the ABCG1 transporter eliminates oxysterol ligands of LXR β , leading to reduced activity of cholesterol-efflux pathways. Internalized and newly synthesized cholesterol can now be used for membrane biogenesis, leading to cell proliferation.

other cell types. Bensinger and colleagues show that LXR activation also inhibits B-cell expansion, whereas LXR-dependent gene expression is reduced when proliferation of another cell type, called fibroblasts, is stimulated. Whether these findings indicate general roles for LXRs in limiting proliferation by promoting cholesterol efflux remains unclear. It will be essential to determine both the relationship between SULT2B1 and ABCG1 expression in other types of rapidly proliferating cell and the consequences of deleting the gene encoding SULT2B1.

Exactly which molecules mediate the checkpoint function of LXR is also unknown. Bensinger *et al.* propose that regulation of cholesterol availability itself is sensed by the cell-cycle machinery. Natural cholesterol precursors and metabolites can activate LXR, but they are probably not sensed by the cell-cycle machinery, as synthetic LXR activators block proliferation.

Although induced LXR activation is sufficient to inhibit T-cell proliferation, it remains possible that SULT2B1 induction affects the SREBP pathway as well as the LXR-mediated pathways. Indeed, several oxysterol substrates of SULT2B1 can inhibit cholesterol biosynthesis¹, possibly by inhibiting the processing of SREBPs to their active form⁸. Thus, SULT2B1-mediated depletion of oxysterols could simultaneously decrease LXR-dependent cholesterol efflux and enhance the activity of the SREBP pathway, as was observed by Bensinger and colleagues. Defining the specific oxysterol species that have regulatory roles in this context will help resolve this issue.

As for T cells, it will be of interest to define the mechanisms that connect the signalling pathways activated by T-cell receptors in response to antigens to the LXR and SREBP pathways, thereby releasing 'the sterol checkpoint' and allowing T-cell expansion. Understanding the basis for SULT2B1 induction may be the key here because of its potential involvement in the reciprocal effects on the LXR and SREBP pathways. Moreover, as T-cell activation also induces the expression of SREBP messenger RNAs and thus its synthesis, T-cell-receptor signalling could have a broader impact on genes that control cholesterol homeostasis.

Might cholesterol-efflux pathways be exploited for therapeutic purposes? Bensinger and colleagues³ note that several 'immortalized' cell lines, including Jurkat T cells, were insensitive to the antiproliferative actions of synthetic LXR ligands, perhaps because they have lost their sensitivity to LXR. This could dim prospects that these compounds might be effective in treating cancers such as leukaemia and lymphomas. The unexpected link between cholesterol homeostasis and adaptive immune responses is worthy of further exploration, however, as suppression of antigen-driven autoimmune diseases by targeting cholesterol efflux and synthesis pathways remains an attractive possibility. ■

Christopher K. Glass and Kaoru Saijo are in the Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093-0651, USA.

e-mails: ckg@ucsd.edu; ksaijo@ucsd.edu

1. Brown, M. S. & Goldstein, J. L. *J. Biol. Chem.* **249**, 7306–7314 (1974).
2. Chen, H. W., Heiniger, H. J. & Kandutsch, A. A. *Proc. Natl Acad. Sci. USA* **72**, 1950–1954 (1975).

3. Bensinger, S. J. *et al. Cell* **134**, 97–111 (2008).
4. Brown, M. S. & Goldstein, J. L. *Cell* **89**, 331–340 (1997).
5. Janowski, B. A., Willy, P. J., Devi, T. R., Falck, J. R. & Mangelsdorf, D. J. *Nature* **383**, 728–731 (1996).
6. Bensinger, S. J. & Tontonoz, P. *Nature* **454**, 470–477 (2008).
7. Chen, W., Chen, G., Head, D. L., Mangelsdorf, D. J. & Russel, D. W. *Cell Metab.* **5**, 73–79 (2007).
8. Radhakrishnan, A., Ikeda, Y., Kwon, H. J., Brown, M. S. & Goldstein, J. L. *Proc. Natl Acad. Sci. USA* **104**, 6511–6518 (2007).

EXPERIMENTAL PHYSICS

A shift in spectroscopy

Frank K. Wilhelm

Spectroscopic measurement of the energy absorbed or emitted by an object is an invaluable experimental technique. An innovative approach opens the door to the acquisition of previously inaccessible data.

For researchers across the sciences, spectroscopy is the main tool for uncovering the energetic structure of their object of study. These data in turn provide a range of information about the object. The idea is that some quantity or other can be characterized as a function of the frequency of radiation that the object absorbs or emits. Typically, the objects of interest, be they nuclei of complex molecules for study by nuclear magnetic resonance or electrons of unknown substances for optical spectroscopy, are probed by weak electromagnetic radiation over a range of frequencies. The object's response — its absorption of the signal or creation of new radiation — usually leads to a sequence of peaks in frequency (ν) called a line spectrum. The lines reveal energetic information through Planck's formula $\Delta E = h\nu$, in which the separation between two of the system's energies (ΔE) is related to ν by

Planck's constant (h). Work by Berns *et al.*¹, described on page 51 of this issue, shows how this principle can be extended: data gained from information provided by the amplitude of the probe radiation can be used to map large parts of an energy spectrum without changing the radiation frequency.

Conventional frequency spectroscopy is a hugely successful technique, but it has its blind spots. First, the frequency and the energy scale of interest are linked by a constant of nature that cannot be changed. In consequence, some energies cannot be probed because radiation of the appropriate frequency, for example frequencies in the terahertz range, are difficult to generate and detect. Second, it may be difficult to deliver the probe radiation to a sample — for instance a cryogenic sample — kept in a protected environment.

The technique of amplitude spectroscopy, as

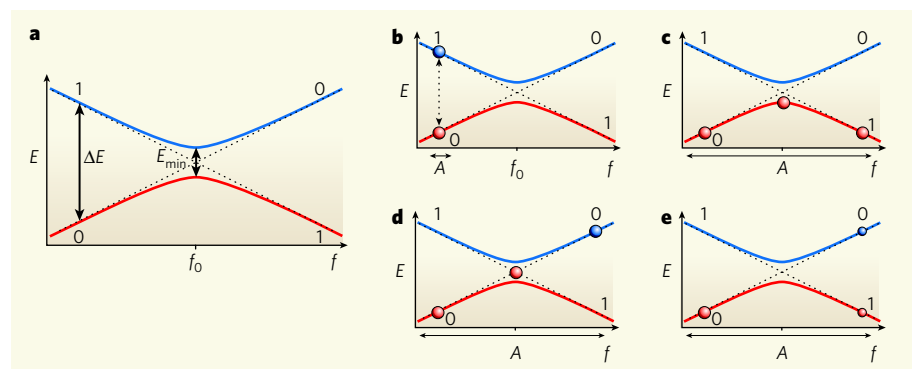


Figure 1 | Use of avoided energy-level crossing for amplitude spectroscopy. **a**, The energies of two quantum states 0 and 1 (dashed lines) approach each other as an applied external field f is changed. Quantum coupling makes the energy levels avoid each other by an energy difference ΔE that is never smaller than E_{\min} , reached at f_0 . The ground state (red line) smoothly crosses over to the excited state (blue line) and vice versa. **b**, Conventional frequency spectroscopy. A signal modulation of small amplitude A induces vertical transitions between the two states (filled circles). **c**, A modulation of large amplitude encompassing f_0 at low frequency leaves the system in the ground state as it changes between 0 and 1 across the avoided crossing. **d**, The same modulation at high frequency leaves the system in state 0, thus crossing over from the ground to the excited state. **e**, The same modulation at intermediate frequency superimposes **c** and **d** and splits the state into two branches. This is the mode of operation in amplitude spectroscopy¹ that makes it possible to track the energy spectrum.

other cell types. Bensinger and colleagues show that LXR activation also inhibits B-cell expansion, whereas LXR-dependent gene expression is reduced when proliferation of another cell type, called fibroblasts, is stimulated. Whether these findings indicate general roles for LXRs in limiting proliferation by promoting cholesterol efflux remains unclear. It will be essential to determine both the relationship between SULT2B1 and ABCG1 expression in other types of rapidly proliferating cell and the consequences of deleting the gene encoding SULT2B1.

Exactly which molecules mediate the checkpoint function of LXR is also unknown. Bensinger *et al.* propose that regulation of cholesterol availability itself is sensed by the cell-cycle machinery. Natural cholesterol precursors and metabolites can activate LXR, but they are probably not sensed by the cell-cycle machinery, as synthetic LXR activators block proliferation.

Although induced LXR activation is sufficient to inhibit T-cell proliferation, it remains possible that SULT2B1 induction affects the SREBP pathway as well as the LXR-mediated pathways. Indeed, several oxysterol substrates of SULT2B1 can inhibit cholesterol biosynthesis¹, possibly by inhibiting the processing of SREBPs to their active form⁸. Thus, SULT2B1-mediated depletion of oxysterols could simultaneously decrease LXR-dependent cholesterol efflux and enhance the activity of the SREBP pathway, as was observed by Bensinger and colleagues. Defining the specific oxysterol species that have regulatory roles in this context will help resolve this issue.

As for T cells, it will be of interest to define the mechanisms that connect the signalling pathways activated by T-cell receptors in response to antigens to the LXR and SREBP pathways, thereby releasing 'the sterol checkpoint' and allowing T-cell expansion. Understanding the basis for SULT2B1 induction may be the key here because of its potential involvement in the reciprocal effects on the LXR and SREBP pathways. Moreover, as T-cell activation also induces the expression of SREBP messenger RNAs and thus its synthesis, T-cell-receptor signalling could have a broader impact on genes that control cholesterol homeostasis.

Might cholesterol-efflux pathways be exploited for therapeutic purposes? Bensinger and colleagues³ note that several 'immortalized' cell lines, including Jurkat T cells, were insensitive to the antiproliferative actions of synthetic LXR ligands, perhaps because they have lost their sensitivity to LXR. This could dim prospects that these compounds might be effective in treating cancers such as leukaemia and lymphomas. The unexpected link between cholesterol homeostasis and adaptive immune responses is worthy of further exploration, however, as suppression of antigen-driven autoimmune diseases by targeting cholesterol efflux and synthesis pathways remains an attractive possibility. ■

Christopher K. Glass and Kaoru Saijo are in the Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093-0651, USA.

e-mails: ckg@ucsd.edu; ksaijo@ucsd.edu

1. Brown, M. S. & Goldstein, J. L. *J. Biol. Chem.* **249**, 7306–7314 (1974).
2. Chen, H. W., Heiniger, H. J. & Kandutsch, A. A. *Proc. Natl Acad. Sci. USA* **72**, 1950–1954 (1975).

3. Bensinger, S. J. *et al. Cell* **134**, 97–111 (2008).
4. Brown, M. S. & Goldstein, J. L. *Cell* **89**, 331–340 (1997).
5. Janowski, B. A., Willy, P. J., Devi, T. R., Falck, J. R. & Mangelsdorf, D. J. *Nature* **383**, 728–731 (1996).
6. Bensinger, S. J. & Tontonoz, P. *Nature* **454**, 470–477 (2008).
7. Chen, W., Chen, G., Head, D. L., Mangelsdorf, D. J. & Russel, D. W. *Cell Metab.* **5**, 73–79 (2007).
8. Radhakrishnan, A., Ikeda, Y., Kwon, H. J., Brown, M. S. & Goldstein, J. L. *Proc. Natl Acad. Sci. USA* **104**, 6511–6518 (2007).

EXPERIMENTAL PHYSICS

A shift in spectroscopy

Frank K. Wilhelm

Spectroscopic measurement of the energy absorbed or emitted by an object is an invaluable experimental technique. An innovative approach opens the door to the acquisition of previously inaccessible data.

For researchers across the sciences, spectroscopy is the main tool for uncovering the energetic structure of their object of study. These data in turn provide a range of information about the object. The idea is that some quantity or other can be characterized as a function of the frequency of radiation that the object absorbs or emits. Typically, the objects of interest, be they nuclei of complex molecules for study by nuclear magnetic resonance or electrons of unknown substances for optical spectroscopy, are probed by weak electromagnetic radiation over a range of frequencies. The object's response — its absorption of the signal or creation of new radiation — usually leads to a sequence of peaks in frequency (ν) called a line spectrum. The lines reveal energetic information through Planck's formula $\Delta E = h\nu$, in which the separation between two of the system's energies (ΔE) is related to ν by

Planck's constant (h). Work by Berns *et al.*¹, described on page 51 of this issue, shows how this principle can be extended: data gained from information provided by the amplitude of the probe radiation can be used to map large parts of an energy spectrum without changing the radiation frequency.

Conventional frequency spectroscopy is a hugely successful technique, but it has its blind spots. First, the frequency and the energy scale of interest are linked by a constant of nature that cannot be changed. In consequence, some energies cannot be probed because radiation of the appropriate frequency, for example frequencies in the terahertz range, are difficult to generate and detect. Second, it may be difficult to deliver the probe radiation to a sample — for instance a cryogenic sample — kept in a protected environment.

The technique of amplitude spectroscopy, as

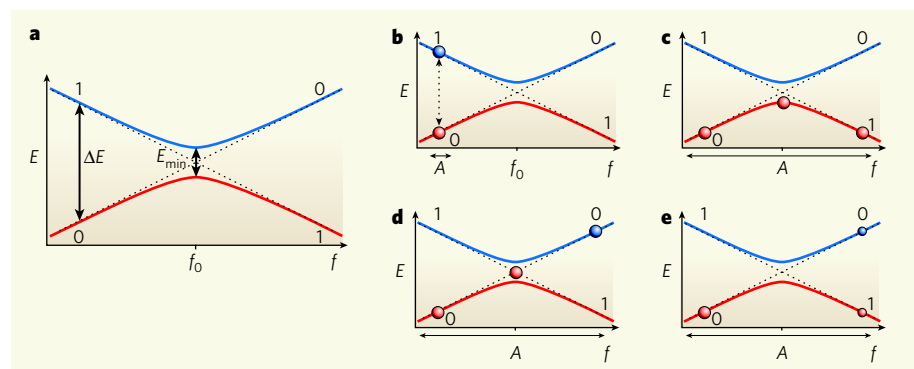


Figure 1 | Use of avoided energy-level crossing for amplitude spectroscopy. **a**, The energies of two quantum states 0 and 1 (dashed lines) approach each other as an applied external field f is changed. Quantum coupling makes the energy levels avoid each other by an energy difference ΔE that is never smaller than E_{\min} , reached at f_0 . The ground state (red line) smoothly crosses over to the excited state (blue line) and vice versa. **b**, Conventional frequency spectroscopy. A signal modulation of small amplitude A induces vertical transitions between the two states (filled circles). **c**, A modulation of large amplitude encompassing f_0 at low frequency leaves the system in the ground state as it changes between 0 and 1 across the avoided crossing. **d**, The same modulation at high frequency leaves the system in state 0, thus crossing over from the ground to the excited state. **e**, The same modulation at intermediate frequency superimposes **c** and **d** and splits the state into two branches. This is the mode of operation in amplitude spectroscopy¹ that makes it possible to track the energy spectrum.

introduced by Berns *et al.*, offers a solution to the problem of reaching energies that correspond to these inaccessible frequencies. A central principle in the authors' approach is quantum interference at an 'avoided' energy-level crossing. An avoided crossing is a quantum phenomenon that has an analogy in mechanics: if two pendulums of identical frequency of oscillation are coupled, they will show new patterns of motion — in phase and in anti-phase — that will have differing frequencies depending on the strength of the coupling. In quantum physics, frequencies correspond to energies, and by the same token, energy levels that are brought close to each other do not cross but keep a minimum distance apart — they show an avoided crossing (Fig. 1a). The quantum physics involved at an avoided crossing is described by the Landau–Zener–Stückelberg (LZS) mechanism^{2–4}. To explain this further, however, some comparisons are needed.

Suppose two quantum energy levels 0 and 1, whose energy is controlled by an external field f , should cross at some value f_0 (Fig. 1a). However, quantum coupling between the levels keeps the energies above a minimum distance apart (E_{\min}). Thus, the energy levels avoid crossing each other. Instead, at around f_0 the ground state is neither 0 nor 1; it is a quantum superposition of both, and the same holds true for the excited state. These superpositions smoothly connect 0 and 1 at around f_0 . In frequency spectroscopy, ν is matched to E/h to map out this energy structure, keeping the signal amplitude (A) as small as possible (Fig. 1b).

The LZS formulation describes large signal amplitudes that sweep across the avoided crossing (Fig. 1c–e). The main parameter is the rate of change of energy $V = \hbar\nu A$, a quantity of dimension E^2 . If V is small, the evolution is adiabatic (that is, no heat enters or leaves), and the system remains in the ground state as it changes between 0 and 1 across the avoided crossing (Fig. 1c). If V is large, the state has no time to change, and remains where it started, in 0 or 1, thus crossing over from the ground to the excited state (Fig. 1d). At intermediate values of V , comparable to E_{\min}^2 , the state splits into a quantum superposition between the two energy branches, creating a wealth of interference patterns (Fig. 1e). The situation in which $V = \hbar\nu A$ is comparable to E_{\min}^2 shows the principle of the frequency conversion: $\hbar\nu$ is multiplied by the large amplitude A to match E_{\min}^2 , allowing A to be traded for frequency.

To track the data for analysis, the energy spectrum is resolved by interferometry. Using detailed knowledge of the interference phenomena generated by a test device, Berns *et al.*¹ reconstruct the full energy spectrum of the device. When an avoided crossing is reached, crossing over to an adjacent state is possible, leading to the 'spectroscopy diamonds' seen in Figure 1a of their paper (page 52)¹. The diamonds, bounded by pairs of avoided crossings that can be located on the f axis from the size of the diamond, contain a periodic interference pattern from which the energy levels away from

the avoided crossing can be extracted. Finally, use of short spectroscopy pulses allows the precise identification of E_{\min} , providing access to previously inaccessible parts of the spectrum.

The experimental test chosen by Berns *et al.* was carried out on a well-controlled macroscopic quantum system known as a superconducting flux qubit⁵. This consists of a small loop of superconducting material interrupted by three Josephson tunnel junctions. Its physics is analogous to that of a particle in a double-well potential whose coordinate is the magnetic flux through the loop, which also serves as the control parameter f . This artificial, engineered device leads to the clean, precise data presented by Berns and colleagues. An advantage of this system for demonstrating amplitude spectroscopy is that because it is macroscopic, its magnetic moment is orders of magnitude larger than that of an atom, making it possible to drive the required large amplitude.

In principle, the amplitude-spectroscopy

scheme described by Berns *et al.*¹ will be widely applicable. But the system must satisfy two requirements: its spectrum must connect the energy levels by avoided crossings; and it must be possible to make long sweeps with the probe radiation without damaging the sample. These demands mean that certain atomic gases and molecular magnets are the most likely candidates for use in such an approach. Amplitude spectroscopy will not replace frequency spectroscopy, but it will complement that technique to complete the picture that researchers can extract from their samples. ■

Frank K. Wilhelm is at the Institute for Quantum Computing, University of Waterloo, Ontario N2L 3G1, Canada.

e-mail: fwilhelm@iqc.ca

1. Berns, D. M. *et al.* *Nature* **455**, 51–57 (2008).
2. Landau, L. D. *Phys. Z. USSR* **2**, 46–51 (1932).
3. Zener, C. *Proc. R. Soc. Lond. A* **137**, 696–702 (1932).
4. Stückelberg, E. C. G. *Helv. Phys. Acta* **5**, 369–422 (1932).
5. Clarke, J. & Wilhelm, F. K. *Nature* **453**, 1031–1042 (2008).

MEDICAL IMAGING

Less is more

Klaas P. Pruessmann

The magnetic resonance imagers used in medicine fill rooms with their large-field magnets. But developments in ultra-low-field devices may give the doctor of tomorrow a more portable version.

Magnetic resonance imaging (MRI) systems are used extensively in the radiology departments of most hospitals. Recent years have seen impressive advances in the quality of the images that MRI produces, in part owing to the use of ever stronger magnetic fields. However, the large, usually cylindrical magnets into which patients are placed are bulky and claustrophobic, leaving little room for patients, let alone much possibility for doctors and researchers to attend to them while in the machine (Fig. 1, overleaf). These magnets are also expensive and heavy, making MRI systems immobile and demanding to install. But all this could change. Writing in the *Journal of Magnetic Resonance*, Vadim Zotev and colleagues¹ report success in imaging a human brain using a different type of MRI system: lightweight, open, mobile and significantly cheaper.

The massive, high-field magnets that are the most obvious feature of MRI machines are responsible for producing a background magnetic field, which has two distinct purposes. First, it polarizes the sample or region of the patient under observation, aligning the magnetic moments of any atomic nuclei that carry a spin. To produce strong polarization, the magnetic field must be large, although it does not need to be particularly homogeneous. Next, the atomic nuclei are excited, usually by an electromagnetic pulse that causes their spins

to flip away from the magnetic-field axis. As a result, the nuclear spins and their associated magnetic moments oscillate and emit electromagnetic signals at a frequency determined by the local magnetic field. To sustain these signals, the background field must be extremely homogeneous. Its strength is less important.

Conventional MRI machines reconcile these different requirements by using magnets that are both powerful and homogeneous. But could the same effect be achieved by using two simpler magnets and switching between them? The first magnet, strong but relatively inhomogeneous, would polarize the sample, whereas the second, weak but highly homogeneous, would be optimized for collecting resonance signals. This concept, termed pre-polarized MRI, was originally introduced by Macovski and Conolly² some 15 years ago, and has been pursued by several research teams since.

Zotev *et al.*¹ now report obtaining images of a living human brain using pre-polarization at 30 millitesla (mT) and image data collection at just 46 μ T, a similar strength to that of Earth's magnetic field and about 30,000 times weaker than that of typical clinical MRI machines. Using such small magnetic fields means that the frequencies of the signals produced by the oscillating nuclear spins are similarly reduced from the usual radiofrequency range to around 2 kilohertz — a frequency readily audible to

introduced by Berns *et al.*, offers a solution to the problem of reaching energies that correspond to these inaccessible frequencies. A central principle in the authors' approach is quantum interference at an 'avoided' energy-level crossing. An avoided crossing is a quantum phenomenon that has an analogy in mechanics: if two pendulums of identical frequency of oscillation are coupled, they will show new patterns of motion — in phase and in anti-phase — that will have differing frequencies depending on the strength of the coupling. In quantum physics, frequencies correspond to energies, and by the same token, energy levels that are brought close to each other do not cross but keep a minimum distance apart — they show an avoided crossing (Fig. 1a). The quantum physics involved at an avoided crossing is described by the Landau–Zener–Stückelberg (LZS) mechanism^{2–4}. To explain this further, however, some comparisons are needed.

Suppose two quantum energy levels 0 and 1, whose energy is controlled by an external field f , should cross at some value f_0 (Fig. 1a). However, quantum coupling between the levels keeps the energies above a minimum distance apart (E_{\min}). Thus, the energy levels avoid crossing each other. Instead, at around f_0 the ground state is neither 0 nor 1; it is a quantum superposition of both, and the same holds true for the excited state. These superpositions smoothly connect 0 and 1 at around f_0 . In frequency spectroscopy, ν is matched to E/h to map out this energy structure, keeping the signal amplitude (A) as small as possible (Fig. 1b).

The LZS formulation describes large signal amplitudes that sweep across the avoided crossing (Fig. 1c–e). The main parameter is the rate of change of energy $V = \hbar\nu A$, a quantity of dimension E^2 . If V is small, the evolution is adiabatic (that is, no heat enters or leaves), and the system remains in the ground state as it changes between 0 and 1 across the avoided crossing (Fig. 1c). If V is large, the state has no time to change, and remains where it started, in 0 or 1, thus crossing over from the ground to the excited state (Fig. 1d). At intermediate values of V , comparable to E_{\min}^2 , the state splits into a quantum superposition between the two energy branches, creating a wealth of interference patterns (Fig. 1e). The situation in which $V = \hbar\nu A$ is comparable to E_{\min}^2 shows the principle of the frequency conversion: $\hbar\nu$ is multiplied by the large amplitude A to match E_{\min}^2 , allowing A to be traded for frequency.

To track the data for analysis, the energy spectrum is resolved by interferometry. Using detailed knowledge of the interference phenomena generated by a test device, Berns *et al.*¹ reconstruct the full energy spectrum of the device. When an avoided crossing is reached, crossing over to an adjacent state is possible, leading to the 'spectroscopy diamonds' seen in Figure 1a of their paper (page 52)¹. The diamonds, bounded by pairs of avoided crossings that can be located on the f axis from the size of the diamond, contain a periodic interference pattern from which the energy levels away from

the avoided crossing can be extracted. Finally, use of short spectroscopy pulses allows the precise identification of E_{\min} , providing access to previously inaccessible parts of the spectrum.

The experimental test chosen by Berns *et al.* was carried out on a well-controlled macroscopic quantum system known as a superconducting flux qubit⁵. This consists of a small loop of superconducting material interrupted by three Josephson tunnel junctions. Its physics is analogous to that of a particle in a double-well potential whose coordinate is the magnetic flux through the loop, which also serves as the control parameter f . This artificial, engineered device leads to the clean, precise data presented by Berns and colleagues. An advantage of this system for demonstrating amplitude spectroscopy is that because it is macroscopic, its magnetic moment is orders of magnitude larger than that of an atom, making it possible to drive the required large amplitude.

In principle, the amplitude-spectroscopy

scheme described by Berns *et al.*¹ will be widely applicable. But the system must satisfy two requirements: its spectrum must connect the energy levels by avoided crossings; and it must be possible to make long sweeps with the probe radiation without damaging the sample. These demands mean that certain atomic gases and molecular magnets are the most likely candidates for use in such an approach. Amplitude spectroscopy will not replace frequency spectroscopy, but it will complement that technique to complete the picture that researchers can extract from their samples. ■

Frank K. Wilhelm is at the Institute for Quantum Computing, University of Waterloo, Ontario N2L 3G1, Canada.

e-mail: fwilhelm@iqc.ca

1. Berns, D. M. *et al.* *Nature* **455**, 51–57 (2008).
2. Landau, L. D. *Phys. Z. USSR* **2**, 46–51 (1932).
3. Zener, C. *Proc. R. Soc. Lond. A* **137**, 696–702 (1932).
4. Stückelberg, E. C. G. *Helv. Phys. Acta* **5**, 369–422 (1932).
5. Clarke, J. & Wilhelm, F. K. *Nature* **453**, 1031–1042 (2008).

MEDICAL IMAGING

Less is more

Klaas P. Pruessmann

The magnetic resonance imagers used in medicine fill rooms with their large-field magnets. But developments in ultra-low-field devices may give the doctor of tomorrow a more portable version.

Magnetic resonance imaging (MRI) systems are used extensively in the radiology departments of most hospitals. Recent years have seen impressive advances in the quality of the images that MRI produces, in part owing to the use of ever stronger magnetic fields. However, the large, usually cylindrical magnets into which patients are placed are bulky and claustrophobic, leaving little room for patients, let alone much possibility for doctors and researchers to attend to them while in the machine (Fig. 1, overleaf). These magnets are also expensive and heavy, making MRI systems immobile and demanding to install. But all this could change. Writing in the *Journal of Magnetic Resonance*, Vadim Zotev and colleagues¹ report success in imaging a human brain using a different type of MRI system: lightweight, open, mobile and significantly cheaper.

The massive, high-field magnets that are the most obvious feature of MRI machines are responsible for producing a background magnetic field, which has two distinct purposes. First, it polarizes the sample or region of the patient under observation, aligning the magnetic moments of any atomic nuclei that carry a spin. To produce strong polarization, the magnetic field must be large, although it does not need to be particularly homogeneous. Next, the atomic nuclei are excited, usually by an electromagnetic pulse that causes their spins

to flip away from the magnetic-field axis. As a result, the nuclear spins and their associated magnetic moments oscillate and emit electromagnetic signals at a frequency determined by the local magnetic field. To sustain these signals, the background field must be extremely homogeneous. Its strength is less important.

Conventional MRI machines reconcile these different requirements by using magnets that are both powerful and homogeneous. But could the same effect be achieved by using two simpler magnets and switching between them? The first magnet, strong but relatively inhomogeneous, would polarize the sample, whereas the second, weak but highly homogeneous, would be optimized for collecting resonance signals. This concept, termed pre-polarized MRI, was originally introduced by Macovski and Conolly² some 15 years ago, and has been pursued by several research teams since.

Zotev *et al.*¹ now report obtaining images of a living human brain using pre-polarization at 30 millitesla (mT) and image data collection at just 46 μ T, a similar strength to that of Earth's magnetic field and about 30,000 times weaker than that of typical clinical MRI machines. Using such small magnetic fields means that the frequencies of the signals produced by the oscillating nuclear spins are similarly reduced from the usual radiofrequency range to around 2 kilohertz — a frequency readily audible to

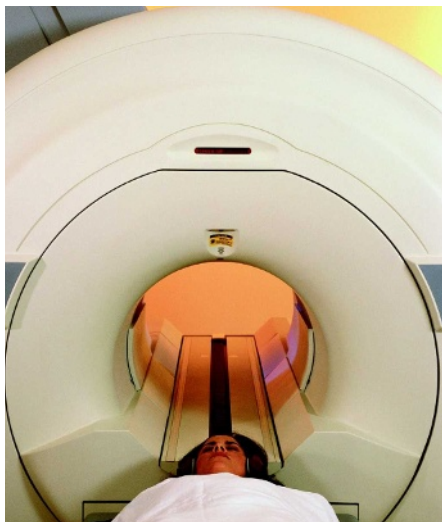


Figure 1 | Tunnel vision. The imposing magnets used to create large, homogeneous fields make magnetic resonance imagers expensive, unwieldy and inconvenient. By dividing the functions of these large-field magnets between two sets of magnets with different characteristics, Zotev *et al.*¹ have produced the prototype of a machine that would be smaller and more open, as well as being capable of performing magnetic resonance imaging and magnetoencephalography at the same time.

the human ear (approximately three octaves above middle C).

Magnetic resonance signals of such low frequency cannot be efficiently detected by conventional means. In fact, when they oscillate so slowly, the extremely weak magnetic moments of the nuclei become hard to detect at all. To observe them, Zotev *et al.*¹ turned to superconducting quantum-interference devices (SQUIDs), which had previously been used for pre-polarized MRI of inanimate samples³. SQUIDs rank among the most sensitive magnetometers created thus far, resolving fields up to the range of femtoTeslas (fT): a billion times smaller than the already small detection field. This sensitivity is sufficient not only to detect nuclear spins in MRI, but also to sense the faint magnetic fields that the brain produces by its neuronal activity, recording of which is known as magnetoencephalography (MEG). Using their low-field approach, Zotev *et al.* have jointly demonstrated MRI and MEG with the same apparatus.

Combined structural MRI and functional MEG would have many applications, but the technology is limited by the speed and overall sensitivity of the MRI part. Zotev *et al.* needed 15 minutes for a full brain scan, and multiple scans had to be averaged to reduce noise in the data. One possible way to improve the performance of microtesla MRI is through detector technology. Advanced devices called atomic magnetometers, which rely on the detection of spin-polarized atoms, have already been proved capable of detecting magnetic resonance⁴ and MEG⁵ signals, and may eventually offer yet higher sensitivity than SQUIDs for both

purposes. Alternatively, large arrays of detectors could be used. Zotev *et al.*¹ used seven SQUIDs simultaneously, but believe that this number could be increased to several hundred. Large detector arrays that cover the whole head would speed up MRI data acquisition by exploiting the different spatial perspectives of each element⁶. In addition, as long as significant noise originates within the detectors, redundant sampling with a dense array will reduce the need to average data collected at different times.

There is also scope for increasing the polarization used in the process. The present results were obtained with a pre-polarizing field of 30 mT, which produced only 2% of the nuclear polarization available in clinical MRI systems. Zotev *et al.*¹ estimate that this field could be approximately tripled. The size of the pre-polarizing field will still be limited, however, by the need to switch the polarizing magnet on and off suitably fast, and by the desire to keep the equipment light and open.

A more radical alternative would be to remove the process of polarization from the patient altogether. Several mechanisms exist to boost nuclear polarization by many orders of magnitude *in vitro*. Hyperpolarized material can then be introduced to the body as the source for an MRI signal. For example, the human lung has already been extensively studied by means

of hyperpolarized ³He gas that patients breathe for the imaging process^{7,8}. At present, there are high hopes for a mechanism known as dynamic nuclear polarization. This permits the hyperpolarization of biomolecules in aqueous solution⁹ that can be injected into the bloodstream. Such mechanisms promise to form a perfect complement to microtesla MRI.

Only time will tell whether affordable, generously open, low-field MRI-MEG systems will one day come into common use in neuroscience labs and radiology departments, but Zotev *et al.* have taken an important step in that direction. ■

Klaas P. Pruessmann is at the Institute for Biomedical Engineering, ETH and University of Zurich, 8092 Zurich, Switzerland.
e-mail: pruessmann@biomed.ee.ethz.ch

1. Zotev, V. S. *et al.* *J. Magn. Reson.* **194**, 115–120 (2008).
2. Macovski, A. & Conolly, S. *Magn. Reson. Med.* **30**, 221–230 (1993).
3. McDermott, R. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 7857–7861 (2004).
4. Savukov, I. M. & Romalis, M. V. *Phys. Rev. Lett.* **94**, 123001 (2005).
5. Xia, H., Ben-Amar Baranga, A., Hoffman, D. & Romalis, M. V. *Appl. Phys. Lett.* **89**, 211104 (2006).
6. Zotev, V. S. *et al.* *J. Magn. Reson.* **192**, 197–208 (2008).
7. van Beek, E. J. R. *et al.* *J. Magn. Reson. Imaging* **20**, 540–554 (2004).
8. Mair, R. W. *et al.* *Magn. Reson. Med.* **53**, 745–749 (2005).
9. Ardenkjær-Larsen, J. H. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 10158–10163 (2003).

SMALL RNAS

The seeds of silence

Zissimos Mourelatos

Individual microRNA sequences can suppress the production of hundreds of proteins. Reduction of protein levels in this way is often modest, however, and many such RNAs probably collectively fine-tune gene expression.

MicroRNAs (miRNAs) are RNA sequences, roughly 23 nucleotides long, that are crucial regulators of gene expression. As part of an RNA-protein complex, these sequences form complementary base pairs with their target messenger RNA sequences, mediating mRNA degradation and/or repressing the translation of the mRNA into protein^{1,2}. Pertinent questions are which and how many proteins a specific miRNA affects, and how it does so. In this issue, Selbach *et al.*³ and Baek *et al.*⁴ address these questions through elegant, large-scale, quantitative studies of the effects of miRNAs on the human and mouse proteomes — the entire set of proteins expressed by the genome.

Over the past five years, large-scale bioinformatics and biochemical approaches have led to the discovery of many miRNA targets. What's more, bioinformatics, coupled with mutational analysis, has uncovered principles of target recognition by miRNAs^{5–7}. For example, one of the main ways in which miRNAs interact with their target mRNAs is through

'seed sites' — that is, continuous base-pairing between nucleotides 2–7 or 8 of a miRNA with a corresponding sequence in the 'miRNA recognition element' (MRE) of the target mRNA. In many cases, the seed seems to determine target recognition single-handedly. In other cases, additional determinants are required, such as more extensive base-pairing between the miRNA and the MRE sequence of the mRNA, and accessibility of this element to the miRNA-protein complex.

When the same (conserved) seed sites are found in related target mRNAs, it is easier to computationally predict MREs. Although many such predicted miRNA targets have been tested, and mRNA profiling has been used as a surrogate for identifying miRNA targets, comprehensive proteomic approaches to measure the effects of miRNAs at the protein level have been lacking — mainly because of the difficulty in accurately comparing and quantifying the effect of miRNAs on the proteome.

In addition to using mRNA profiling to

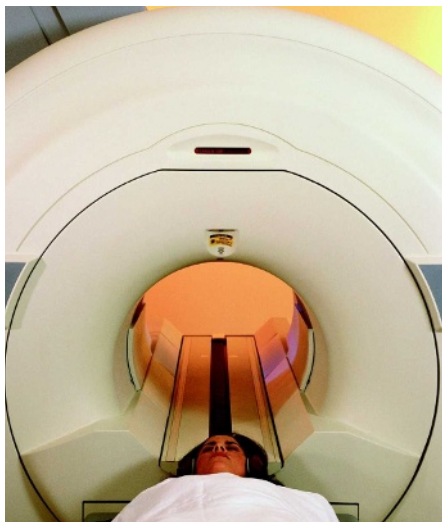


Figure 1 | Tunnel vision. The imposing magnets used to create large, homogeneous fields make magnetic resonance imagers expensive, unwieldy and inconvenient. By dividing the functions of these large-field magnets between two sets of magnets with different characteristics, Zotev *et al.*¹ have produced the prototype of a machine that would be smaller and more open, as well as being capable of performing magnetic resonance imaging and magnetoencephalography at the same time.

the human ear (approximately three octaves above middle C).

Magnetic resonance signals of such low frequency cannot be efficiently detected by conventional means. In fact, when they oscillate so slowly, the extremely weak magnetic moments of the nuclei become hard to detect at all. To observe them, Zotev *et al.*¹ turned to superconducting quantum-interference devices (SQUIDs), which had previously been used for pre-polarized MRI of inanimate samples³. SQUIDs rank among the most sensitive magnetometers created thus far, resolving fields up to the range of femtoTeslas (fT): a billion times smaller than the already small detection field. This sensitivity is sufficient not only to detect nuclear spins in MRI, but also to sense the faint magnetic fields that the brain produces by its neuronal activity, recording of which is known as magnetoencephalography (MEG). Using their low-field approach, Zotev *et al.* have jointly demonstrated MRI and MEG with the same apparatus.

Combined structural MRI and functional MEG would have many applications, but the technology is limited by the speed and overall sensitivity of the MRI part. Zotev *et al.* needed 15 minutes for a full brain scan, and multiple scans had to be averaged to reduce noise in the data. One possible way to improve the performance of microtesla MRI is through detector technology. Advanced devices called atomic magnetometers, which rely on the detection of spin-polarized atoms, have already been proved capable of detecting magnetic resonance⁴ and MEG⁵ signals, and may eventually offer yet higher sensitivity than SQUIDs for both

purposes. Alternatively, large arrays of detectors could be used. Zotev *et al.*¹ used seven SQUIDs simultaneously, but believe that this number could be increased to several hundred. Large detector arrays that cover the whole head would speed up MRI data acquisition by exploiting the different spatial perspectives of each element⁶. In addition, as long as significant noise originates within the detectors, redundant sampling with a dense array will reduce the need to average data collected at different times.

There is also scope for increasing the polarization used in the process. The present results were obtained with a pre-polarizing field of 30 mT, which produced only 2% of the nuclear polarization available in clinical MRI systems. Zotev *et al.*¹ estimate that this field could be approximately tripled. The size of the pre-polarizing field will still be limited, however, by the need to switch the polarizing magnet on and off suitably fast, and by the desire to keep the equipment light and open.

A more radical alternative would be to remove the process of polarization from the patient altogether. Several mechanisms exist to boost nuclear polarization by many orders of magnitude *in vitro*. Hyperpolarized material can then be introduced to the body as the source for an MRI signal. For example, the human lung has already been extensively studied by means

of hyperpolarized ³He gas that patients breathe for the imaging process^{7,8}. At present, there are high hopes for a mechanism known as dynamic nuclear polarization. This permits the hyperpolarization of biomolecules in aqueous solution⁹ that can be injected into the bloodstream. Such mechanisms promise to form a perfect complement to microtesla MRI.

Only time will tell whether affordable, generously open, low-field MRI-MEG systems will one day come into common use in neuroscience labs and radiology departments, but Zotev *et al.* have taken an important step in that direction.

Klaas P. Pruessmann is at the Institute for Biomedical Engineering, ETH and University of Zurich, 8092 Zurich, Switzerland.
e-mail: pruessmann@biomed.ee.ethz.ch

1. Zotev, V. S. *et al.* *J. Magn. Reson.* **194**, 115–120 (2008).
2. Macovski, A. & Conolly, S. *Magn. Reson. Med.* **30**, 221–230 (1993).
3. McDermott, R. *et al.* *Proc. Natl Acad. Sci. USA* **101**, 7857–7861 (2004).
4. Savukov, I. M. & Romalis, M. V. *Phys. Rev. Lett.* **94**, 123001 (2005).
5. Xia, H., Ben-Amar Baranga, A., Hoffman, D. & Romalis, M. V. *Appl. Phys. Lett.* **89**, 211104 (2006).
6. Zotev, V. S. *et al.* *J. Magn. Reson.* **192**, 197–208 (2008).
7. van Beek, E. J. R. *et al.* *J. Magn. Reson. Imaging* **20**, 540–554 (2004).
8. Mair, R. W. *et al.* *Magn. Reson. Med.* **53**, 745–749 (2005).
9. Ardenkjær-Larsen, J. H. *et al.* *Proc. Natl Acad. Sci. USA* **100**, 10158–10163 (2003).

SMALL RNAS

The seeds of silence

Zissimos Mourelatos

Individual microRNA sequences can suppress the production of hundreds of proteins. Reduction of protein levels in this way is often modest, however, and many such RNAs probably collectively fine-tune gene expression.

MicroRNAs (miRNAs) are RNA sequences, roughly 23 nucleotides long, that are crucial regulators of gene expression. As part of an RNA-protein complex, these sequences form complementary base pairs with their target messenger RNA sequences, mediating mRNA degradation and/or repressing the translation of the mRNA into protein^{1,2}. Pertinent questions are which and how many proteins a specific miRNA affects, and how it does so. In this issue, Selbach *et al.*³ and Baek *et al.*⁴ address these questions through elegant, large-scale, quantitative studies of the effects of miRNAs on the human and mouse proteomes — the entire set of proteins expressed by the genome.

Over the past five years, large-scale bioinformatics and biochemical approaches have led to the discovery of many miRNA targets. What's more, bioinformatics, coupled with mutational analysis, has uncovered principles of target recognition by miRNAs^{5–7}. For example, one of the main ways in which miRNAs interact with their target mRNAs is through

'seed sites' — that is, continuous base-pairing between nucleotides 2–7 or 8 of a miRNA with a corresponding sequence in the 'miRNA recognition element' (MRE) of the target mRNA. In many cases, the seed seems to determine target recognition single-handedly. In other cases, additional determinants are required, such as more extensive base-pairing between the miRNA and the MRE sequence of the mRNA, and accessibility of this element to the miRNA-protein complex.

When the same (conserved) seed sites are found in related target mRNAs, it is easier to computationally predict MREs. Although many such predicted miRNA targets have been tested, and mRNA profiling has been used as a surrogate for identifying miRNA targets, comprehensive proteomic approaches to measure the effects of miRNAs at the protein level have been lacking — mainly because of the difficulty in accurately comparing and quantifying the effect of miRNAs on the proteome.

In addition to using mRNA profiling to

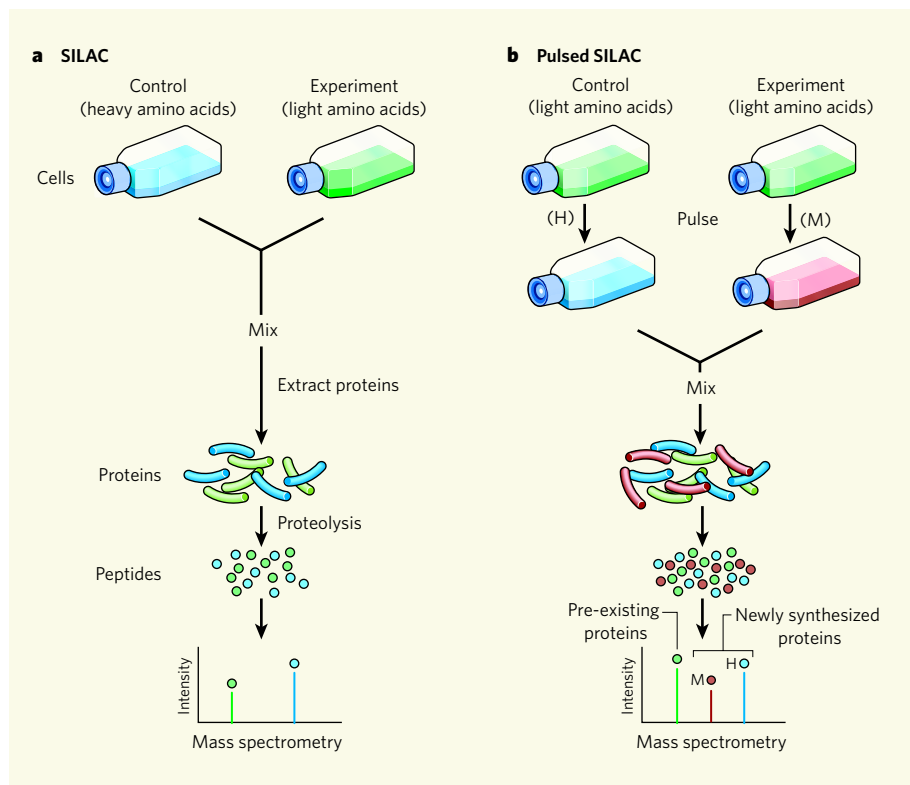


Figure 1 | The techniques of SILAC and pulsed SILAC. **a**, Baek *et al.*⁴ used the standard SILAC technique. Here, cells are grown in culture media such that one set of cells, for example untreated cells (control), is exposed to essential amino acids containing the heavy form of an isotope, whereas another set, treated cells (experiment), is exposed to the same amino acids labelled with the light form of the isotope. The cells are then mixed, and following extraction and enzymatic proteolysis of their proteins, the peptides generated are analysed by mass spectrometry. Peptides corresponding to the same proteins from the two samples can then be differentiated based on their labelling, and quantified. **b**, Selbach *et al.*³ used a modified version of SILAC called pulsed SILAC. In this method, pulsed exposure of cells (control and experiment) already treated with amino acids labelled with two different heavy versions of an isotope — medium-heavy (M) or heavy (H) — allows identification and quantification of only newly synthesized proteins.

measure the effects of miRNAs at the mRNA level, Selbach *et al.*³ and Baek *et al.*⁴ use a powerful mass spectrometric method called SILAC ('stable-isotope labelling with amino acids in cell culture'), which allows changes in protein levels to be monitored in treated and control cells⁸ (Fig. 1a). Baek and colleagues (page 64) separately introduced three miRNAs — miR-124, miR-1 and miR-181 — into the human HeLa cell line to study the effect of these regulatory sequences on the expression of proteins in the nucleus. In another set of experiments, they investigated the consequences of miRNA deficiency by analysing the proteome of neutrophil cells derived from mice lacking miR-223.

As expected, the expression of certain proteins was reduced following the introduction of miR-124, miR-1 and miR-181, and that of others was increased in neutrophils lacking miR-223. The authors then searched the mRNAs affected — both in regions that were translated into protein and those that were not — looking for specific nucleotide motifs. Their analysis revealed that the motifs predominantly found in the mRNA for affected proteins were seeds located in the untranslated regions at one end (the 3' end) of the mRNA, which corresponded

to the miRNA sequence that affected them.

In most cases, the changes in protein levels were modest (about 1.5–2 fold), and were paralleled by concomitant changes in the mRNA levels. Intriguingly, the expression of many proteins whose mRNA contained non-conserved seeds for miR-223 was increased in the absence of this miRNA, suggesting that they are also miR-223 targets. Prediction of non-conserved sites is difficult, but Baek and colleagues show that ranking such sites by 'total context score', which includes site type (such as complementarity with nucleotide 8), number of sites and site context (such as sites embedded in adenine–uracil-rich areas of the 3' untranslated region), could be an effective way to identify functional sites.

Selbach *et al.*³ (page 58) either separately introduced five miRNAs (miR-1, miR-155, miR-16, miR-30a and let-7b) or reduced the expression of let-7b in HeLa cells. They then analysed the effects of these manipulations by a deft modification of SILAC called pulsed SILAC, which allows analysis of only newly synthesized proteins following a treatment (Fig. 1b). The authors' main conclusions parallel those of Baek *et al.*, in that the seed sites

seem to be the only motifs that best correlate with miRNA-induced changes in the proteome. The presence of several seeds and the conservation of seeds seem to result in stronger miRNA effects, but once again the overall changes in protein levels for the miRNAs tested were modest.

Selbach and colleagues find that, although many miRNA targets are affected by both repression of translation and mRNA degradation, a number of targets are predominantly regulated by translational repression. Many of these latter targets correspond to mRNAs that are translated on ribosomes associated with the cellular organelle known as the endoplasmic reticulum, as opposed to ribosomes free in the cytoplasm. So at least for a subset of miRNA targets, translational repression may indeed be the first step in miRNA-mediated regulation of gene expression, followed by mRNA degradation in the longer term.

The studies' findings^{3,4} greatly enhance our understanding of how miRNAs recognize many of their targets, and provide a genome-wide view of the effects of miRNAs in regulating gene expression. They also pave the way to the study of miRNAs by quantitative proteomics. But not all miRNA-induced changes in the proteome correlate with the presence of seeds (conserved or non-conserved) in the mRNAs of the affected proteins^{3,4}. Some of these changes may be secondary, but others are probably the result of direct targeting by miRNAs. Moreover, the false-positive rate of target predictions, even with seed incorporation, is an estimated 40%.

A challenge will therefore be to elucidate, in even greater detail, how miRNAs recognize their targets, and to identify the factors that modulate target recognition and miRNA activity. Are there common structural features for sites that do not rely on seeds? Do miRNA levels influence the number of targeted mRNAs? How does the structure of the mRNA influence recognition by the miRNA? Are there MRE sequences, or certain structural features of these sequences, that promote translational repression instead of mRNA degradation? Many of these questions can now be addressed using the elegant proteomics approaches described in these studies^{3,4}.

Zissimos Mourelatos is in the Department of Pathology and Laboratory Medicine, Division of Neuropathology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

e-mail: mourelaz@uphs.upenn.edu

1. Liu, X., Fortin, K. & Mourelatos, Z. *Brain Pathol.* **18**, 113–121 (2008).
2. Du, T. & Zamore, P. D. *Development* **132**, 4645–4652 (2005).
3. Selbach, M. *et al.* *Nature* **455**, 58–63 (2008).
4. Baek, D. *et al.* *Nature* **455**, 64–71 (2008).
5. Bartel, D. P. & Chen, C. Z. *Nature Rev. Genet.* **5**, 396–400 (2004).
6. Rajewsky, N. *Nature Genet.* **38** (Suppl.), S8–S13 (2006).
7. Sethupathy, P., Megraw, M. & Hatzigeorgiou, A. G. *Nature Methods* **3**, 881–886 (2006).
8. Mann, M. *Nature Rev. Mol. Cell Biol.* **7**, 952–958 (2006).

OBITUARY

Victor Almon McKusick (1921–2008)

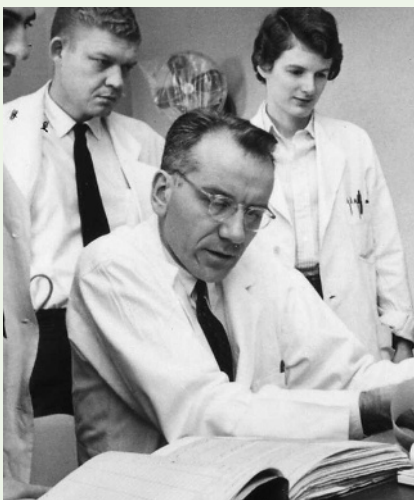
Quiet revolutionary in genetic medicine.

Following the complete sequencing of the human genome, we stand at the beginning of an era that promises medical treatments tailored to an individual's genetic make-up. No one is more responsible for this revolution than Victor McKusick, who died on 22 July. McKusick was the first to understand that systematically mapping human genes predisposing the bearer to disease, which many considered no better than stamp-collecting, was a route to a new medicine. In this and other ways he was instrumental in moulding the discipline that we now call genetic medicine, and in making genetics the basic science of medicine.

McKusick was born on a dairy farm in Maine in 1921. His early ambition was to enter the ministry. At the age of 15, however, a streptococcal infection of his arm that required a long hospital stay and treatment with one of the first antibiotics, made him rethink his future. His identical twin Vincent chose to study law. Victor, by contrast, after initial education at Tufts University, entered Johns Hopkins University in 1943 to pursue medical training, making a name for himself in cardiology.

Like medicine, genetics came to him by chance. His fascination with one teenage patient who suffered from intestinal polyps and melanin spots, and later with three members of a family who exhibited the same syndrome, provided him with first-hand experience of the basic principles of genetics. One was the need to recognize patterns of inheritance, in this case dominant as opposed to recessive, that suggested mutations at one genetic location. Another was the need to distinguish between mechanisms: in these patients, were two genes involved, one for polyps and one for spots, which were co-inherited (linkage), or were polyps and spots different manifestations of the same gene (pleiotropy)? McKusick was thus well armed when he subsequently came across patients with Marfan syndrome — with its dominant inheritance and remarkable pleiotropy affecting the aorta, eye and skeleton — which, he argued, arose from mutations in a single gene.

Similar patients and their families were to prove pivotal in his conversion to genetics, which was complete by 1957. Asked to direct a chronic-disease clinic by his boss, McKusick argued that “genetic disease is the ultimate chronic disease, since it's lifelong”, and seized the opportunity to reshape the Moore Clinic at Johns Hopkins to create the first unit devoted to medical genetics. He learnt his trade by doing: by using the rudimentary cytogenetic, biochemical and



population (quantitative) genetic methods then available. He soon became convinced of three guiding principles: the value of knowing a gene's location in the human genome; the value of accumulated genetic information; and the value of disseminating this new information widely and rapidly.

Given the individual rarity of most hereditary disorders, McKusick knew that he had to learn about the experiences of others and to share his own. He was a prolific organizer, of both ideas and facts, a trait most notably made manifest in 1966 in *Mendelian Inheritance in Man* (*MIM*), the first edition of his catalogue of all known genes and genetic disorders. The final print edition appeared in 1998, but since 1987 it has also been available as full-text online, with a free database (www.ncbi.nlm.nih.gov/Omim). It now has some 19,000 entries, with more than 70% of the content having been produced by McKusick himself. This is his most lasting achievement — it is a deep resource and knowledge base, without which clinicians and any manner of biologist would be intellectually orphaned.

One of McKusick's preoccupations was with cataloguing the location of each human gene associated with a disease, and thus to create a disease map of the human genome. He did this not only through his own pioneering studies, but by beginning — chiefly with Frank Ruddle — a series of human gene-mapping workshops. Subsequently, he was an influential voice in organizing the international community around the Human Genome Organization (HUGO, fondly called Victor's HUGO). For him, the *raison d'être* of mapping, which he articulated in 1969, well before anyone understood or believed it, was that mapping all human genes was the best way to understand the basic malfunctions causing birth defects.

The existence of *MIM*, together with McKusick's mapping preoccupation, were the two most persuasive factors in favour of the public project to sequence the human genome. McKusick himself was on the US National Research Council committee that recommended the project, and was one of its prime cheer-leaders. He was among those who argued for a ‘map first, sequence later’ approach, and was a supporter of mapping and sequencing other species, and of tackling the whole genome rather than only the known functional genome.

As a pragmatist, however, McKusick was also attracted to Craig Venter's idea of sequencing expressed sequence tags (nucleic-acid snippets that encode only a portion of functional genes). He supported both the public sequencing project and Venter's private sequencing effort at Celera (he was a trustee of Venter's eponymous institute), because he believed that the genome could thus be completed sooner. The leaders of both the public and private sequencing ventures (Francis Collins and Venter, respectively) paid their respects at his funeral service.

McKusick made research on the human species, despite its poor genetic properties of few offspring and long generation times, a treasure trove for uncovering new genetic mechanisms. He also provided a glimpse of the future for genetic medicine in an interview given in 2001: “I think the medical geneticist will spend much more time overseeing gene screens, or genome screens, interpreting the results to individuals, and designing programs to make the most of the strong points of the genome and to avoid troubles from some of the weak points in the genome.” Spreading the word was a vital part of his legacy — as for example in the influential ‘Short Course’ in mammalian genetics, held annually at the Jackson Laboratory in Bar Harbor, Maine, which he founded in 1960 and co-directed.

In the long journey to his many accomplishments, Victor McKusick was accompanied by his rheumatologist wife Anne. Those accomplishments are all the more remarkable for having been achieved without his once raising his voice. But then a man who had genetics institutes named after him in Baltimore, Bologna and Beijing had no need to draw attention to himself.

Aravinda Chakravarti

Aravinda Chakravarti is at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland 21093, USA. e-mail: aravinda@jhmi.edu

JOHNS HOPKINS MEDICINE

The future of biocuration

To thrive, the field that links biologists and their data urgently needs structure, recognition and support.

Doug Howe, Seung Yon Rhee *et al.*

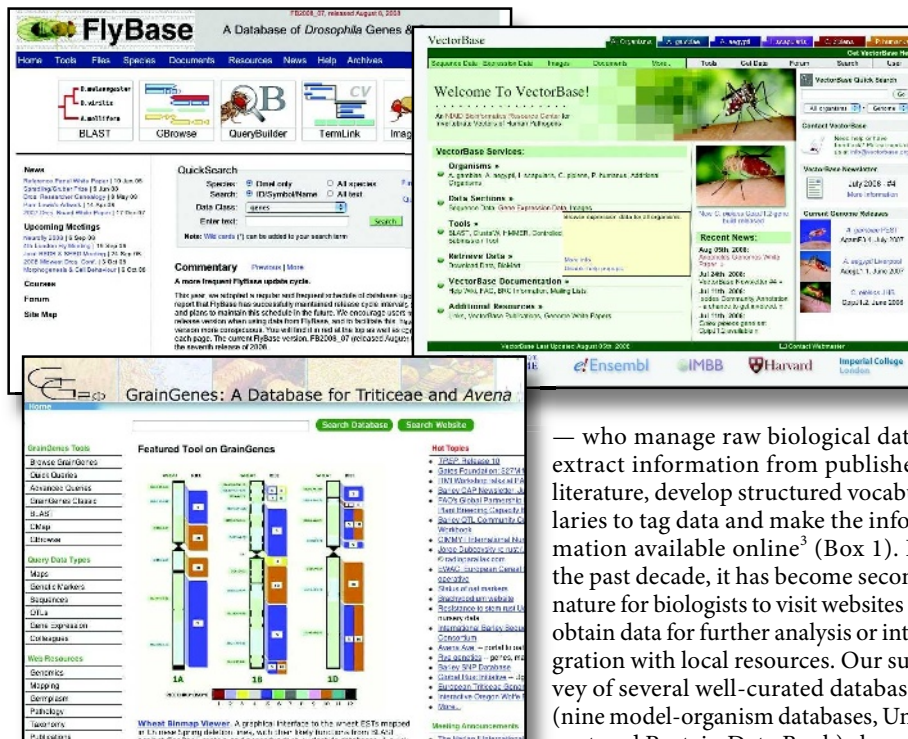


The exponential growth in the amount of biological data means that revolutionary measures are needed for data management, analysis and accessibility. Online databases have become important avenues for publishing biological data. Biocuration, the activity of organizing, representing and making biological information accessible to both humans and computers, has become an essential part of biological discovery and biomedical research. But curation increasingly lags behind data generation in funding, development and recognition.

We propose three urgent actions to advance this key field. First, authors, journals and curators should immediately begin to work together to facilitate the exchange of data between journal publications and databases. Second, in the next five years, curators, researchers and university administrations should develop an accepted recognition structure to facilitate community-based curation efforts. Third, curators, researchers, academic institutions and funding agencies should, in the next ten years, increase the visibility and support of scientific curation as a professional career.

Failure to address these three issues will cause the available curated data to lag farther behind current biological knowledge. Researchers will observe an increasing occurrence of obvious gaps in knowledge. As these gaps expand, resources will become less effective for generating and testing hypotheses, and the usefulness of curated data will be seriously compromised.

When all the data produced or published are curated to a high standard and made accessible as soon as they become available, biological research will be conducted in a manner that is quite unlike the way it is done now. Researchers will be able to process massive amounts of complex data much more quickly. They will garner insight about the areas of their interest rapidly with the help of inference programs. Digesting information and generating hypotheses at the computer screen will be so much faster that researchers will get back to the bench quickly for more experiments. Experiments will be designed with more insight; this increased specificity will cause an exponential growth in



— who manage raw biological data, extract information from published literature, develop structured vocabularies to tag data and make the information available online³ (Box 1). In the past decade, it has become second nature for biologists to visit websites to obtain data for further analysis or integration with local resources. Our survey of several well-curated databases (nine model-organism databases, UniProt and Protein Data Bank) showed that nearly 750,000 visitors (unique IP addresses) viewed more than 20 million pages in just one month (March 2008, Eva Huala, Peter Rose, Rolf Apweiler, personal communications).

Despite the essential part that it plays in today's research, biocuration has been slow to develop. To provide a forum for the exchange of ideas and methods, and to facilitate collaborations and training, more than 150 biocurators met at two international conferences and created a mailing list and a website (www.biocurator.org). These meetings and discussions have honed in on the three actions, outlined above and elaborated on below, that must now be addressed to ensure scientists' continued access to the high-quality data on which their research depends.

Come together

Extracting, tagging with controlled vocabularies, and representing data from the literature, are some of the most important and time-consuming tasks in biocuration. Curated information from the literature serves as the gold-standard data set for computational analysis, quality assessment of high-throughput data and benchmarking of data-mining

knowledge, much as we are experiencing exponential growth in data today.

Data avalanche

Biology, like most scientific disciplines, is in an era of accelerated information accrual and scientists increasingly depend on the availability of each others' data. Large-scale sequencing centres, high-throughput analytical facilities and individual laboratories produce vast amounts of data such as nucleotide and protein sequences, protein crystal structures, gene-expression measurements, protein and genetic interactions and phenotype studies. By July 2008, more than 18 million articles had been indexed in PubMed and nucleotide sequences from more than 260,000 organisms had been submitted to GenBank^{1,2}. The recently announced project to sequence 1,000 human genomes in three years to reveal DNA polymorphisms (www.1000genomes.org) is a tip of the data iceberg.

Such data, produced at great effort and expense, are only as useful as researchers' ability to locate, integrate and access them. In recent years, this challenge has been met by a growing cadre of biologists — 'biocurators'

algorithms. Meanwhile, the boundaries of the biological domain that researchers study are widening rapidly, so researchers need faster and more reliable ways to understand unfamiliar domains. This too is facilitated by literature curation.

Typically, biocurators read the full text of articles and transfer the essence into a database. For a paper about the molecular biology of a particular gene, process or pathway, such information might include gene-expression patterns, mutant phenotypes, results of biochemical assays, protein-complex membership and the authors' inferences about the functions and roles of the gene products studied. As each paper uses different experimental and analysis methods, capturing this information in a consistent fashion requires intensive thought and effort. Limited resources and staff mean that most curation groups can't keep up with all the relevant literature.

How information is presented in the literature greatly affects how fast biocurators can identify and curate it. Papers still often report newly cloned genes without providing GenBank IDs or the species from which the genes were cloned. The entities discussed in a paper, including species, genes, proteins, genotypes and phenotypes must be unambiguously identified during curation. For example, using the HUGO Gene Nomenclature Committee resource (www.genenames.org), we find that the human gene *CDKN2A* has ten literature-based synonyms. One of those, *p14*, is also a synonym for five other genes: *CDK2AP2*, *CTNBL1*, *RPP14*, *S100A9* and *SUB1*. To confirm the identity of the gene described, curators make inferences from synonyms, reported sequences, biological context and bibliographic citations. This time-consuming and error-prone step could be eliminated by compliance with data-reporting standards⁴⁻⁹.

Most recent efforts in this direction have been developed by the communities that produce large-scale genomics data. The vast majority of the peer-reviewed literature does not yet have a reporting-structure standard. As publication has become a mainly digital endeavour, however, publications and biological databases are becoming increasingly similar. Properly cross-referenced and indexed, each could serve as an access point to the other¹⁰. Such collaboration between databases and journals would improve researchers' access to data and make their work more visible.

We recommend that all journals and reviewers require that a distinct section of the Methods (or a supplemental document) of all published articles includes approved gene symbols (which are inherently unstable) and model-organism database IDs (which do not change) for genes discussed; nucleotide or protein accession numbers (GenBank or UniProt ID) for isoforms of each gene or protein

Box 1 | The role of biocurators

- To extract knowledge from published papers
- To connect information from different sources in a coherent and comprehensible way
- To inspect and correct automatically predicted gene structures and protein sequences to provide high-quality proteomes
- To develop and manage structured controlled vocabularies that are crucial for data relations and the logical retrieval of large data sets
- To integrate knowledge bases to represent complex systems such as metabolic pathways and protein-interaction networks.
- To correct inconsistencies and errors in data representation
- To help data users to render their research more productive in a timely manner
- To steer the design of web-based resources
- To interact with researchers to facilitate direct data submissions to databases

discussed; and descriptions of species, strains, cell types and genotypes used. Examples of sources for this information are listed in Table 1. This would accelerate literature curation, uphold information integrity, facilitate the proper linkage of data to other resources and support automated mining of data from papers. Another model is for authors to provide a 'structured digital abstract' — a machine-readable XML summary of pertinent facts in the article¹¹ — along with a manuscript. This approach is in an experimental phase at the journal *FEBS Letters*¹².

Journals should also mandate direct submission of data into appropriate databases as a part of publication. This has been implemented by the journal *Plant Physiology* and curators of The *Arabidopsis* Information Resource (TAIR) database¹³. On acceptance of a manuscript, the corresponding author must fill out a simple web-based form to provide appropriate genetic and molecular information about the *Arabidopsis* genes in the publication. The information is sent to TAIR for integration by biocurators, who work with the authors to ensure that the data reported are of high quality and accurate.

As this infrastructure develops, we would like to see authors routinely tagging all aspects of the data in their publication semantically using universally agreed tag standards. Examples of such tags include the National Center for Biotechnology Information (NCBI) Taxon IDs, the Gene Ontology (GO) IDs and Enzyme Commission (EC) numbers. This information should be embedded in the electronic versions of publications or provided in a supplemental file similar to the crystallographic information file (CIF) currently required for publication of a crystal structure. The CIF file is submitted to the Protein Data Bank (www.pdb.org), which

offers software to assist in preparation and validation of such crystallographic data¹⁴. An analogous system to help authors identify, tag and validate the crucial basic information in their research reports before publication would accelerate the automated linkage of literature to key records in existing databases and improve the accuracy of the published data.

In short, authors and publishers must use the existing publication infrastructure to facilitate literature curation much more to the benefit of all parties.

Community curation

Curation of large-scale genomics and post-genomics data enjoys no such luxury of 'an existing publication infrastructure' to leverage, although emerging standards of data reporting are promising⁴⁻⁹. Sooner or later, the research community will need to be involved in the annotation effort to scale up to the rate of data generation. This transition will require annotation tools, standardized methods, oversight by expert curators and a combination of social infrastructure, tool development, training and feedback. Biocurators are especially important for establishing such an infrastructure and training to maintain consistency and accuracy.

To date, not much of the research community is rolling up its sleeves to annotate. What will be the tipping point? The main limitation in community annotation is the perceived lack of incentive. For example, several model-organism databases have requested that authors annotate the genes they publish. This has historically failed for one main reason: contributions by experts consist of information they already know, and do not increase the value of the resource to themselves. A mechanism tied to career or research advancement may be required before community curation can be established as a broadly accepted and productive scientific endeavour¹⁵. Incentives for researchers to curate data should include new information or insight for their research interests, improvement in academic reputation or impact, career advancement and better funding chances. Academic departments and funding agencies should consider community annotation as a productive contribution to the scientific research corpus and a natural extension of the publication process.

For example, in the *Daphnia* Genomics Consortium (<http://daphnia.cgb.indiana.edu>) collaboration wiki, a community of more than 300 contributors took ownership of annotation of the genome while it was being sequenced at the Joint Genome Institute in Walnut Creek, California, and shared publication authorship as a consortium. Similarly, the International *Glossina* Genomics Initiative (<http://iggi.sanbi.ac.za>) hosted an annotation jamboree for field workers, population geneticists and molecular biologists to annotate tsetse fly molecular data as the sequence information became available. This

"To date, not much of the research community is rolling up its sleeves to annotate."

consortium-based publication mechanism is analogous to that used by other large-scale scientific projects such as the Sloan Digital Sky Survey (www.sdss.org). This is a viable course for communities that lack funding for dedicated curators, and offers a reward structure through consortium publication for participation and subsequent satellite papers.

The recently launched WikiProfessional Life Sciences (www.wikiprofessional.org) project links community curation with research and reputation gains. WikiProfessional indexed more than one million authors from PubMed and comparable numbers of biological concepts from authoritative databases and generated a simple way for researchers to update the information¹⁶. Because new potential 'facts' are mined from the network of associated concepts, the more accurate and comprehensive a

particular concept is, the more chance it will have of being associated with other relevant ones, which in turn will lead to more potential new facts. All the updates researchers make are immediately publicly visible under their own name. Similarly, the Gene Wiki project generated thousands of wiki stubs in Wikipedia for human genes in an attempt to make it easier for the community to update the gene pages¹⁷. Although these wiki-based approaches provide an infrastructure for contributors to be recognized, there is not yet a standard practice for these contributions to be cited like a publication. It is imperative that the researchers, journal publishers and database curators start building a standard mechanism for citing annotation data sets.

Allowing anyone with a web browser, including the general public, to annotate

entries would increase the number of potential annotators substantially, as pioneered in several astronomy projects. At Galaxy Zoo (www.galaxyzoo.org), 80,000 astronomers and members of the public manually classified the morphology of one million galaxies in less than three weeks. An analogous system to allow the public to contribute to biological annotation could be just as powerful if presented properly. For example, one could show a user an image of an *in situ* hybridization experiment and ask them to grade it as 'not expressed', 'restricted expression' or 'ubiquitous expression'. Even such basic information, if available for many thousands of genes, would be useful as first-pass annotation.

In sum, researchers (and even the general public) can be mobilized to provide the substantial resources needed to address the immense volume of data, if participation is appropriately rewarded. In the next five years, curators, funding agencies and academic institutions alike must find ways to consider substantial contributions to community curation efforts, much like a peer-reviewed publication, when it comes to issues of promotion, salary, hiring and funding.

Career path

How can biocuration mature faster as a career? Biocurators currently streamline submission to databases, automate curation, standardize data and facilitate contributions to annotation by research communities interested in the annotation process. To handle the increasing volume and types of data, journal publishers and researchers who generate data will need to be involved in the curation process and the roles of biocurators will expand to include editing and teaching. As biology moves towards more precise, quantitative science, biologists also need to adapt to thinking more quantitatively, systematically and objectively about their data; biocuration will need to become an inherent part of research and education in biology.

Biocuration requires a blend of skills and experience, including advanced scientific research and competence in database management systems, multiple operating systems and scripting languages. This type of background has typically been garnered through a combination of self-teaching and on-the-job experience, which can be narrow and spotty. Happily, formal education is becoming available. For example, the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign offers a biological information specialist master's degree and a specialization in data curation¹⁸. Experienced biocurators must lead the way in establishing more and better formal training programmes. In the next 5–10 years, biology curricula should include courses in biocuration as this becomes an increasingly common activity for all biological researchers. And interdisciplinary programmes that include courses in

Table 1 | Examples of knowledge-sharing databases

Species	Database	URL
Model organism databases		
<i>Aedes aegypti</i>	VectorBase	www.vectorbase.org
<i>Anopheles gambiae</i>	VectorBase	www.vectorbase.org
<i>Arabidopsis thaliana</i>	The <i>Arabidopsis</i> Information Resource	www.arabidopsis.org
<i>Caenorhabditis elegans</i>	WormBase	www.wormbase.org
<i>Candida albicans</i>	<i>Candida</i> Genome Database	www.candidagenome.org
<i>Culex pipiens</i>	VectorBase	www.vectorbase.org
<i>Danio rerio</i>	Zebrafish Information Network	http://zfin.org
<i>Dictyostelium discoideum</i>	dictyBase	http://dictybase.org
<i>Drosophila</i> sp.	FlyBase	http://flybase.org
<i>Glycine max</i>	SoyBase	www.soybase.org
<i>Homo sapiens</i>	HUGO Gene Nomenclature Committee	www.genenames.org
<i>Hordeum vulgare</i>	Barley Genetic Stocks Database	http://ace.untamo.net/bgs
<i>Ixodes scapularis</i>	VectorBase	www.vectorbase.org
<i>Leishmania</i> sp.	GeneDB	www.genedb.org
<i>Mus musculus</i>	Mouse Genome Informatics	www.informatics.jax.org
<i>Oryza</i> sp.	Gramene	http://gramene.org
<i>Paramecium tetraurelia</i>	ParameciumDB	http://paramecium.cgm.cnrs-gif.fr
<i>Pediculus humanus</i>	VectorBase	www.vectorbase.org
<i>Rattus norvegicus</i>	Rat Genome Database	http://rgd.mcw.edu
<i>Saccharomyces cerevisiae</i>	Saccharomyces Genome Database	www.yeastgenome.org
<i>Schizosaccharomyces pombe</i>	GeneDB	www.genedb.org
<i>Solanaceae</i> sp.	Sol Genomics Network	http://sgn.cornell.edu
<i>Strongylocentrotus purpuratus</i>	SpBase	http://sugp.caltech.edu/SpBase
<i>Triticum</i> sp.	GrainGenes	http://wheat.pw.usda.gov
<i>Trypanosoma</i> sp.	GeneDB	www.genedb.org
<i>Xenopus laevis</i>	Xenbase	www.xenbase.org
<i>Xenopus tropicalis</i>	Xenbase	www.xenbase.org
<i>Zea mays</i>	Maize Genetics and Genomics Database	www.maizegdb.org
Nucleotide, protein and structure databases		
All Species	GenBank	www.ncbi.nlm.nih.gov/Genbank
All Species	UniProt	www.pir.uniprot.org
All Species	Protein Data Bank	http://rcsb.org/pdb/home/home.do
Taxonomy		
All Species	NCBI Entrez Taxonomy	www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy

Biological databases contain unique identifiers for the unambiguous identification of biological entities (such as genes, proteins, species and chemicals). These identifiers do not change as common biological names do. Authors should consult these databases for stable identifiers to cite in their publications.

biology, computer science and information science will be vital.

Attracting highly qualified individuals into this field has been challenging. The whole community must promote scientific curation as a professional career option. Funding agencies must assess the impact of curated data and support the development of innovative curation methods. To improve the profession, curators need a forum to share their experiences and publish their works. Oxford University Press plans to begin publishing a new journal in 2009 called *Database: The Journal of Biological Databases and Curation*. This may provide one such venue for publication of noteworthy advances in biocuration (www.database.oxfordjournals.org). Meanwhile, a committee of 20 biocurators and researchers is forming an International

Society for Biocuration (www.biocurator.org/BiocuratorSociety.html) to make the discipline more visible and to promote it as an attractive career path. The official launch of the society is planned for the third International Biocuration Meeting next April in Berlin (<http://projects.embl.org/Meeting2009>).

Biology today needs more robust, expressive, computable, quantitative, accurate and precise ways to handle data. It is time to recognize that biocuration and biocurators are central to the future of the field. ■

1. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. *Nucl. Acid. Res.* **36**, D25–D30 (2008).
2. Wheeler, D. L. et al. *Nucl. Acid. Res.* **36**, D13–D21 (2008).
3. Salimi, N. & Vita, R. *PLoS Comput. Biol.* **2**, e125 (2006).
4. Brazma, A. et al. *Nature Genet.* **29**, 365–371 (2001).
5. Deutsch, E. W. et al. *Nature Biotechnol.* **26**, 305–312 (2008).
6. Field, D. et al. *Nature Biotechnol.* **26**, 541–547 (2008).

7. Jenkins, H. et al. *Nature Biotechnol.* **22**, 1601–1606 (2004).
8. Orchard, S. et al. *Nature Biotechnol.* **25**, 894–898 (2007).
9. Taylor, C. F. et al. *Nature Biotechnol.* **25**, 887–893 (2007).
10. Bourne, P. *PLoS Comput. Biol.* **1**, 179–181 (2005).
11. Seringhaus, M. R. & Gerstein, M. B. *BMC Bioinformatics* **8**, 17 (2007).
12. Seringhaus, M. & Gerstein, M. *FEBS Lett.* **582**, 1170 (2008).
13. Ort, D. R. & Grennan, A. K. *Plant Physiol.* **146**, 1022–1023 (2008).
14. Burkhardt, K., Schneider, B. & Ory, J. *PLoS Comput. Biol.* **2**, e99 (2006).
15. Rhee, S. Y. *Plant Physiol.* **134**, 543–547 (2004).
16. Mons, B. et al. *Genome Biol.* **9**, R89 (2008).
17. Huss, J. W. et al. *PLoS Biol.* **6**, e175 (2008).
18. Palmer, C. L., Heidorn, P. B., Wright, D. & Cragin, M. H. *Int. J. Dig. Curation* **2**, 31–40 (2007).

Author information Correspondence and requests for materials should be addressed to D.H. (e-mail: dhowe@cs.uoregon.edu) and S.Y.R. (e-mail: rhee@acoma.stanford.edu).

See Editorial, page 1.

Authorship

Doug Howe¹, Maria Costanzo², Petra Fey³, Takashi Gojobori⁴, Linda Hannick⁵, Winston Hide^{6,7}, David P. Hill⁸, Renate Kania⁹, Mary Schaeffer^{10,11}, Susan St Pierre¹², Simon Twigger¹³, Owen White¹⁴ and Seung Yon Rhee¹⁵

¹The Zebrafish Information Network, 5291 University of Oregon, Eugene, Oregon 97403-5291, USA. ²*Saccharomyces* and *Candida* Genome Databases, Stanford University, Stanford, California 94305-5120, USA. ³dictyBase, Northwestern University Biomedical Informatics Center, 750 N. Lake Shore Drive, 11-175, Chicago, Illinois 60611, USA.

⁴Centre for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization of Information and Systems, Yata, Mishima 411-8540, Japan. ⁵J. Craig Venter Institute, Applied Bioinformatics, Rockville, Maryland 20850, USA. ⁶South African National Bioinformatics Institute, University of the Western Cape, Private Bag X17, Bellville 7535, South Africa. ⁷Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, USA. ⁸Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine 04609, USA. ⁹Scientific Databases and Visualization, EML Research GmbH, Villa Bosch, Schloss-Wolfsbrunnengasse 33, D-69118 Heidelberg, Germany. ¹⁰Division of Plant Sciences, University of Missouri, Columbia, Missouri, USA. ¹¹Plant Genetics Research Unit, Agricultural Research Service, United States Department of Agriculture, Columbia, Missouri 65211-7020, USA. ¹²FlyBase, Harvard University, Cambridge, Massachusetts 02138, USA. ¹³Rat Genome Database, Bioinformatics Research Center, Medical College of Wisconsin, 8701 Watertown Plank Rd, Milwaukee, Wisconsin 53226, USA.

¹⁴Department of Epidemiology and Preventative Medicine, Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.

¹⁵The *Arabidopsis* Information Resource, Carnegie Institution for Science, Department of Plant Biology, 260 Panama Street, Stanford, California 94305, USA.

Amplitude spectroscopy of a solid-state artificial atom

David M. Berns^{1,2}, Mark S. Rudner¹, Sergio O. Valenzuela^{3†}, Karl K. Berggren^{4†}, William D. Oliver^{2,4}, Leonid S. Levitov¹ & Terry P. Orlando^{2,5}

The energy-level structure of a quantum system, which has a fundamental role in its behaviour, can be observed as discrete lines and features in absorption and emission spectra. Conventionally, spectra are measured using frequency spectroscopy, whereby the frequency of a harmonic electromagnetic driving field is tuned into resonance with a particular separation between energy levels. Although this technique has been successfully employed in a variety of physical systems, including natural and artificial atoms and molecules, its application is not universally straightforward and becomes extremely challenging for frequencies in the range of tens to hundreds of gigahertz. Here we introduce a complementary approach, amplitude spectroscopy, whereby a harmonic driving field sweeps an artificial atom through the avoided crossings between energy levels at a fixed frequency. Spectroscopic information is obtained from the amplitude dependence of the system's response, thereby overcoming many of the limitations of a broadband-frequency-based approach. The resulting 'spectroscopy diamonds', the regions in parameter space where transitions between specific pairs of levels can occur, exhibit interference patterns and population inversion that serve to distinguish the atom's spectrum. Amplitude spectroscopy provides a means of manipulating and characterizing systems over an extremely broad bandwidth, using only a single driving frequency that may be orders of magnitude smaller than the energy scales being probed.

Spectroscopy has historically been used to obtain a wide range of information about atomic and nuclear properties^{1,2}. In early work, the determination of spectral lines helped elucidate the principles of quantum mechanics through studies of the hydrogen atom and provided a means of testing atomic theory. Since then, several spectroscopy techniques to determine absolute transition frequencies (or, equivalently, wavelengths) have been developed, involving the emission, absorption or scattering (for example Raman) of radiation. The advent of tuneable, coherent radiation sources at microwave and optical frequencies led to the age of modern atomic spectroscopy, in which a primary approach is to identify absorption spectra of natural^{1,2} and artificial^{3–14} atoms and molecules as the source frequency ν is varied to satisfy the resonance conditions $\Delta E = h\nu$, where ΔE is the energy-level separation and h is Planck's constant.

The study of artificial atoms, whose spectra extend into the microwave and millimetre-wave regimes (10–300 GHz), faces distinct challenges. Stable, tuneable microwave sources of frequencies in excess of 70 GHz exist, but are expensive and generally require multipliers that are inefficient and intrinsically noisy¹⁵. Frequency-dependent dispersion and attenuation, tight tolerances to control impedance, and multi-mode or restricted-bandwidth performance of transmission lines and waveguides¹⁵ limit the application of broadband-frequency spectroscopy in these systems, particularly in cryogenic environments. Despite these challenges, superposition states in superconducting^{3–5} and semiconducting artificial atoms⁷ have been probed directly up to frequencies of several tens of gigahertz. A number of leading groups have developed alternative techniques, for example resonant- and photon-assisted tunnelling^{8,9}, that can be used to access spectroscopic information in specific systems at

even higher frequencies, although each has its own advantages and limitations and may not be easily applicable to other systems.

Amplitude spectroscopy, introduced here, probes the energy-level structure of a quantum system through its response to driving-field amplitude rather than frequency (Fig. 1a). It is applicable to systems with avoided energy-level crossings that can be traversed using an external control parameter, including solid-state artificial atoms^{3–14}, molecular magnets^{16–18} and spin systems¹⁹. Such 'longitudinal' excursions through the energy-level diagram (Fig. 1c) are achieved by strong driving with an external field at a fixed frequency, which may be several orders of magnitude lower than the frequencies required for direct resonance with the varying energy-level spacings. For appropriate combinations of amplitude and frequency, the quantum evolution is adiabatic, except in the vicinity of avoided energy-level crossings where Landau–Zener-type quantum-coherent transitions²⁰ occur. The quantum interference between repeated Landau–Zener transitions gives rise to Stückelberg interference fringes²¹ that encode information about the system's coherent evolution and energy spectrum. By trading amplitude for frequency, the amplitude spectroscopy approach makes it possible to probe and manipulate quantum systems, in particular those with strong coupling to external fields, over wide bandwidths. In our experiment, we determine the energy spectrum of a manifold of states with energies from $h \times 0.01$ GHz to $h \times 120$ GHz in a superconducting artificial atom, using a driving frequency near 0.1 GHz.

Implementation

We demonstrate amplitude spectroscopy with a superconducting qubit, a solid-state artificial atom that has discrete energy states⁸

¹Department of Physics, ²Research Laboratory for Electronics, ³Francis Bitter Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴Lincoln Laboratory, Massachusetts Institute of Technology, 244 Wood Street, Lexington, Massachusetts 02420, USA. ⁵Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present addresses: ICREA and Centre d'Investigacions en Nanociència i Nanotecnologia, UAB Campus, 08193 Bellaterra, Spain (S.O.V.); Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (K.K.B.).

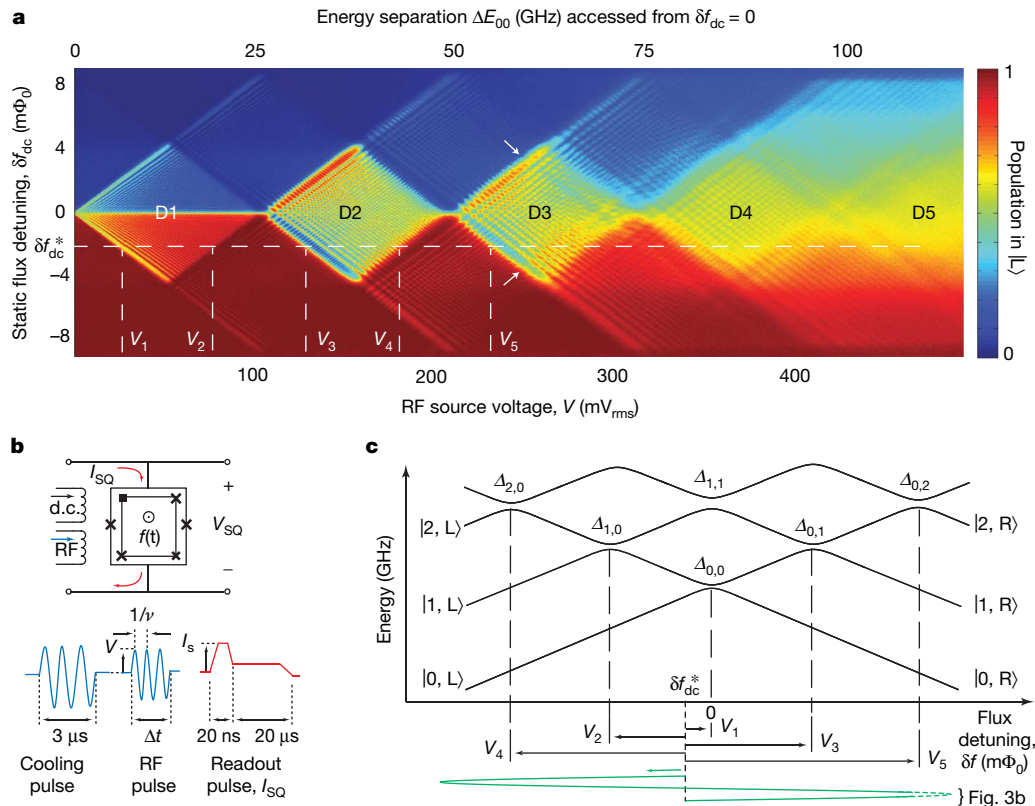


Figure 1 | Amplitude spectroscopy with long-pulse driving towards saturation. **a**, Amplitude spectroscopy diamonds. The qubit is driven at a fixed frequency of $\nu = 0.16$ GHz, and the driving amplitude V is swept for each static flux detuning δf_{dc} . The colour scale indicates net fractional qubit population in state $|L\rangle$ (see text). The main diamond regions, which are symmetric about $\delta f_{dc} = 0$, are labelled D1–D5. Their edges mark the parameter values where particular level crossings are first reached, for example, amplitudes V_1 – V_5 for $\delta f_{dc} = \delta f_{dc}^*$. Arrows indicate signatures of transverse modes in D3 (Fig. 4). The top axis shows the $|0, L\rangle$ – $|0, R\rangle$ energy spacing, $\Delta E_{0,0}$, accessed when driving with an amplitude V from $\delta f_{dc} = 0$.

which can be strongly coupled to external radio-frequency fields while preserving quantum coherence¹⁰. Artificial atoms are natural systems in which to probe a wide range of quantum effects: coherent superpositions of macroscopic states^{3–6}, Rabi oscillations^{10–14,22–25}, incoherent Landau–Zener transitions²⁶, Stückelberg oscillations^{27–30}, microwave cooling^{31–33}, cavity quantum electrodynamics^{34–38} and aspects of quantum measurement^{39–41}.

Our qubit (Fig. 1b) is a niobium superconducting loop interrupted by three Josephson junctions^{42,43} (see Supplementary Information). Near flux bias $f \approx 0.5 \Phi_0$, where the unit Φ_0 is the superconducting flux quantum, the qubit potential has a two-dimensional double-well profile parameterized by the flux detuning $\delta f \equiv f - 0.5 \Phi_0$ (Supplementary Fig. 1a). The qubit potential is approximately separable at lower energies, so the system's first few energy eigenstates can be assigned transverse ($p = 0, 1, 2, \dots$) and longitudinal ($q = 0, 1, 2, \dots$) quantum numbers, with energies controlled by the flux detuning δf . When the potential is tilted so that resonant inter-well tunnelling is suppressed, the eigenstates closely approximate the diabatic well states localized in the left- and right-hand wells, which are associated with loop currents of opposing circulation. In this limit, the energies of localized states in the left- and right-hand wells respectively increase and decrease approximately linearly with flux detuning (Fig. 1c). Whenever the diabatic states $|p, q, L\rangle$ (left well) and $|p', q', R\rangle$ (right well) are degenerate, resonant inter-well tunnelling mixes them and opens avoided crossings $\Delta_{pq,p'q'}$. Because, for an ideally symmetric system, our driving is longitudinal and therefore conserves the parity of the transverse modes, we assume initially that

b, Schematic of the qubit surrounded by a SQUID magnetometer readout. Static (d.c.) and radio-frequency (RF) fields control the state of the qubit: a 3- μ s cooling-pulse³¹ (11 MHz, 990 mV) followed by an amplitude spectroscopy pulse of duration Δt . The qubit state is read out using a SQUID current pulse, I_{SQ} , while monitoring the presence of a SQUID voltage V_{SQ} . **c**, Energy-level diagram illustrating the relation between the driving amplitude V and the level-crossing positions for a particular static flux detuning $\delta f_{dc} = \delta f_{dc}^*$. Arrows represent the amplitudes V_1 – V_5 at which the crossings are reached, marking the edges of the spectroscopic diamonds in **a**.

only the lowest transverse mode is populated, and use the reduced notation $|q, L\rangle$ for $|p, q, L\rangle$, $|q', R\rangle$ for $|p', q', R\rangle$ and $\Delta_{q,q'}$ for $\Delta_{pq,p'q'}$ (Supplementary Fig. 1b). We do observe, however, signatures of weak excitations of transverse modes (see the discussion below).

We drive the qubit longitudinally with a time-dependent flux (Fig. 1c, green sinusoid)

$$\delta f(t) = \delta f_{dc} - \Phi_{rf} \sin \omega t \quad (1)$$

that induces sinusoidal excursions through the energy levels about a static flux bias δf_{dc} , where the driving amplitude $\Phi_{rf} = \alpha V$ is proportional to the source voltage V with a frequency-dependent constant of proportionality α . To reach a regime dominated by Landau–Zener transitions at level crossings, we choose the driving frequency $\nu = \omega/2\pi$ such that $h\nu$ is generally much smaller than the instantaneous energy-level spacing throughout the driving cycle but the evolution through level crossings is non-adiabatic. The transition rate between the states q and q' is controlled by the relative-energy sweep rate

$$\zeta_i = h(|m_q| + |m_{q'}|) \frac{d}{dt} \delta f|_{t=t_i} = h(|m_q| + |m_{q'}|) \Phi_{rf} \omega \cos \omega t_i \quad (2)$$

evaluated at the time t_i at which the system is swept through an avoided crossing $\Delta_{q,q'}$. Here

$$m_q = \frac{1}{h} \frac{dE_q}{df}$$

is the diabatic energy-level slope of state q in units of frequency per

flux. In this regime, a Landau–Zener transition at an avoided crossing with energy splitting $\Delta_{q,q'}$ occurs with probability

$$P_{\text{LZ}} = 1 - \exp\left(-\frac{\pi \Delta_{q,q'}^2}{2\hbar \zeta_i}\right) \quad (3)$$

where $\hbar = h/2\pi$. Such transitions drive the system into a coherent superposition of energy eigenstates associated with different wells.

Repeated Landau–Zener transitions give rise to Stückelberg oscillations^{21,27–30,44} in the populations of the states q and q' . For a crossing $\Delta_{q,q'}$, and using a fixed driving frequency ν , the resulting interference patterns depend on the driving amplitude Φ_{rf} through the sweep rate ζ_i and on the static flux bias δf_{dc} through the times t_i . Analysing the interference patterns in $(\Phi_{\text{rf}}, \delta f_{\text{dc}})$ space, therefore, allows us to obtain spectroscopic information about the system. Because the rate ζ_i is proportional to both amplitude and frequency, we can accommodate a small driving frequency by compensating with a large driving amplitude at an appropriate static flux bias. This also allows us to control the time interval between consecutive Landau–Zener transitions through a given crossing. For Stückelberg interference^{21,27–30,44} to occur, this time interval, which is typically a small fraction of the driving period $1/\nu$, must be smaller than the relevant decoherence times²⁹ (see below).

Each experiment uses the pulse sequence shown in Fig. 1b, which consists of a harmonic cooling pulse to initialize the qubit in its

ground state³¹, followed by the desired amplitude spectroscopy pulse. The qubit state is determined by applying a synchronous readout pulse to a superconducting quantum interference device (SQUID) magnetometer (see Supplementary Information). Using this technique, we investigate both the long-time and short-time behaviour of our qubit, and determine the energy-level slopes m_q along with the splittings $\Delta_{q,q'}$ and d.c.-flux locations $\delta f_{q,q'}$ of level crossings that constitute the energy-level diagram.

Stationary amplitude spectroscopy

Figure 1a displays the amplitude spectroscopy of the qubit driven towards saturation. Four primary spectroscopy diamonds (D1, D2, D3 and D4) with large population contrast, centred about $\delta f_{\text{dc}} = 0$, are observed in the data; they are flanked by eight fainter diamonds. The diamond structures result from the interplay between static flux detuning and driving amplitude, which determine when the various level crossings are reached. Because the onset of each diamond is associated with transitions at a particular level crossing, the boundaries of the diamonds mark the occurrence of level crossings. We use the linear relation between V and Φ_{rf} (Fig. 2a) to obtain the values of $\delta f_{q,q'}$ listed in Table 1.

For the particular static flux detuning $\delta f_{\text{dc}} = \delta f_{\text{dc}}^* < 0$ (Fig. 1a, c, dashed line), the cooling pulse prepares the qubit in the ground state, $|0, L\rangle$. As the amplitude of the spectroscopy pulse is increased from

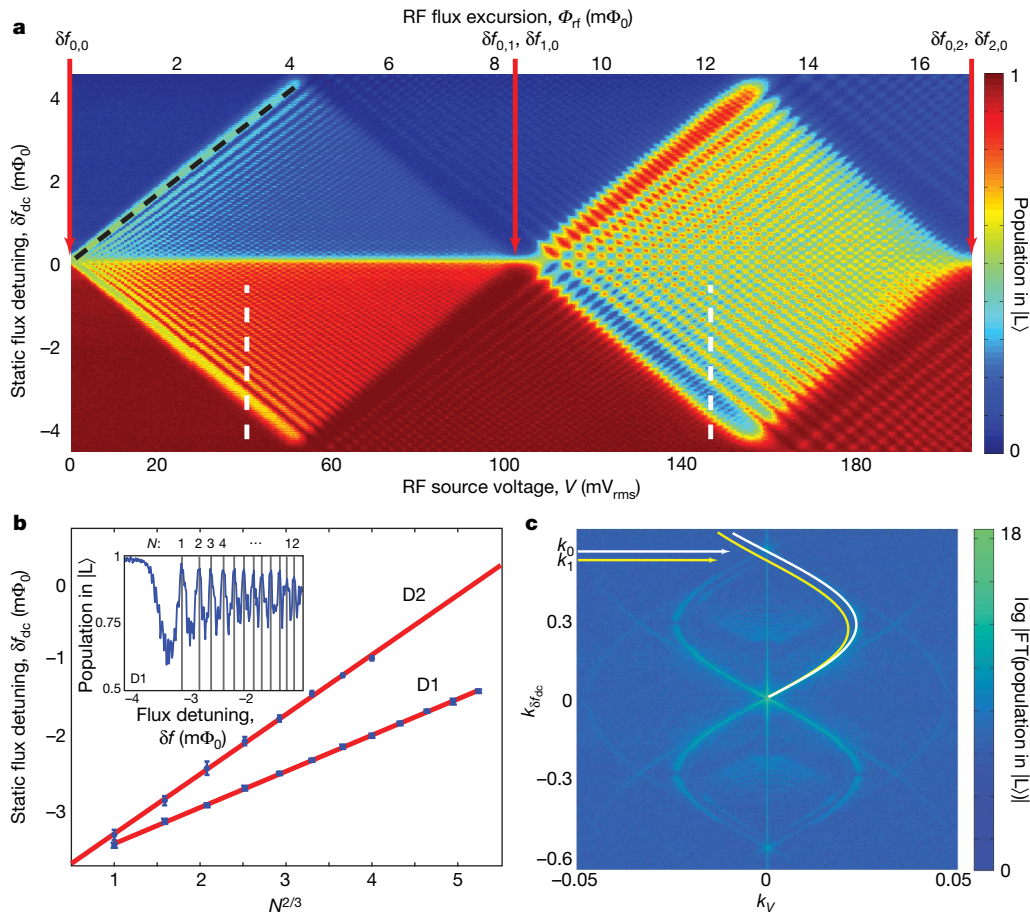


Figure 2 | Energy-level slopes and interference patterns. **a**, Detail of D1 and D2 (Fig. 1a) showing interference patterns due to single (D1) and multiple (D2) avoided crossings (see text). We note the strong population inversion in D2, and the cooling in the region between D1 and D2 as well as at interference nodes in D2. Arrows indicate the locations $\delta f_{q,q'}$ of avoided crossing (top axis); the flux-to-voltage conversion factor α is determined by the left-hand side of D1 (dashed black line). **b**, Determination of the energy-level slopes for levels $|0, L\rangle$, $|0, R\rangle$, $|1, L\rangle$ and $|1, R\rangle$ from the interference fringes (dashed white lines in **a**) at 43 mV_{rms} (D1) and 150 mV_{rms} (D2). We

plot detuning location of the N th interference nodes (see inset) versus $N^{2/3}$, and corresponding linear fits (red lines). The error bars indicate residual estimates from identifying node positions. The inset shows a vertical slice from D1 (43 mV_{rms}); interference-node positions used for scaling analysis are indicated by vertical lines. **c**, Discrete 2DFT of both spectroscopic diamonds in **a**. The sinusoids with half-periods k_0 and k_1 , used to extract the energy-level slopes (see text), arise from crossings $\Delta_{0,0}$ and $\Delta_{1,0}$ ($\Delta_{0,1}$). The reciprocal-space variables $k_{\delta f_{\text{dc}}}$ and k_V correspond to the real-space variables δf_{dc} and V , respectively.

Table 1 | Energy spectrum parameters determined using amplitude spectroscopy

Crossing, q, q'	Location, $\delta f_{a,q'} \text{ (m}\Phi_0\text{)}$	Magnitude, $A_{a,q}/h \text{ (GHz)}$	Energy-level slope, $m_{q'} \text{ (GHz m}\Phi_0^{-1}\text{)}$
0, 0	0	0.013 ± 0.001	1.44 ± 0.01
0, 1	8.4 ± 0.2	0.090 ± 0.005	1.09 ± 0.03
0, 2	17.0 ± 0.2	0.40 ± 0.01	0.75 ± 0.04
0, 3	25.8 ± 0.4	2.2 ± 0.1	0.49 ± 0.08

$V = 0$, population transfer from $|0, L\rangle$ to $|0, R\rangle$ first occurs at $V = V_1$, where the $A_{0,0}$ crossing is reached (Fig. 1a, left-hand side of D1; other level-crossing voltages are similarly labelled V_2, \dots, V_5). For $V_1 < V < V_2$, Stückelberg interference^{27–30,44} at the $A_{0,0}$ crossing results in the observed fringe contrast (Fig. 2a). At $V = V_2$, the adjacent crossing, $A_{1,0}$, is reached (Fig. 1c), inducing transitions between levels $|0, R\rangle$ and $|1, L\rangle$ and marking the right-hand side of D1.

For $V_2 < V < V_3$, the data show a large reduction in contrast due to the addition of a single, strong transition at $A_{1,0}$. The saturated population depends on the competition between transitions at $A_{0,0}$ and $A_{1,0}$, on relatively fast intra-well relaxation and, to a lesser extent, on much slower inter-well relaxation processes. In our qubit, because $A_{0,0} \ll A_{1,0}$, the dominant transitions occur at the $A_{1,0}$ crossing. Transitions $|0, L\rangle \rightarrow |0, R\rangle$ are still induced at the $A_{0,0}$ crossing, but constructive Stückelberg interference at $A_{1,0}$ converts a substantial fraction of that population to $|1, L\rangle$, an excited state of the left-hand well. Because relaxation within a well is a relatively fast process in this qubit in comparison with the relaxation between wells, the excited-state population tends to relax back to the ground state, $|0, L\rangle$, thus suppressing the net population transfer. In contrast, for values of V such that the interference at $A_{1,0}$ is destructive, the population remains in $|0, R\rangle$, making the interference fringes arising from $A_{0,0}$ visible, albeit with reduced contrast (Fig. 1a, faint diamond; Fig. 2a). For $V_2 < V < V_3$, the qubit can be cooled to its ground state³¹; in this work we also initialize the qubit in this regime.

At even larger amplitudes, transitions to additional excited states become possible. For $V > V_3$, the qubit can make transitions between $|0, L\rangle$ and $|1, R\rangle$, marking the left-hand side of D2. The right-hand side of D2 is marked by the amplitude, $V = V_4$, where $A_{2,0}$ is reached, allowing transitions between $|0, R\rangle$ and $|2, L\rangle$. This description can be extended straightforwardly to the remainder of the spectrum. In this qubit, we did not find explicit signatures of coherent multi-path traversal between the $\delta f < 0$ and $\delta f > 0$ regions of the energy-level diagram (for example, through avoided crossings $A_{1,1}$, $A_{2,2}$ and so on).

There are several notable features associated with amplitude spectroscopy. First, we are able to probe the qubit continuously over an extremely broad bandwidth. In particular, spectroscopy diamond D5 (Fig. 1a) results from transitions to energy levels more than $h \times 100$ GHz above the ground state. Even at such high energy levels, our artificial atom retains its energy-level structure in the presence of the strong driving field used to probe it.

Second, we use a single driving frequency of only 0.16 GHz. Generally, for double-well systems, the splittings $A_{q,q'}$ tend to increase in higher excited states (Fig. 1c). In such cases, the entire spectrum can be mapped using a single frequency, or a small range of frequencies, because the larger driving amplitudes required to reach those larger splittings $A_{q,q'}$ also provide the larger sweep rates required to probe them.

Third, diamond D2 shows strong population inversion due to the competition between transitions to the respective excited states $|1, L\rangle$ and $|1, R\rangle$ at avoided crossings $A_{1,0}$ and $A_{0,1}$ combined with fast intra-well relaxation to $|0, L\rangle$ and $|0, R\rangle$ (Fig. 2a). The transition rates at $A_{1,0}$ and $A_{0,1}$ have strong oscillatory behaviour due to Stückelberg interference, which is constructive or destructive depending on the values of δf_{dc} and V . The competition between these rates leads to the observed checkerboard pattern, symmetric about $\delta f_{dc} = 0$, with alternating regions of strong population inversion and efficient cooling, depending on the specific well (left-hand well or right-hand well)

in which the relaxation occurs. Similar checkerboard patterns are present in the diamonds D3 and D4. Because the population inversion observed here relies on relaxation, it loses its coherence with the driving field and can be used as the active medium of a single-atom laser⁴⁵.

The energy-level separation $\Delta E_{q,q'} \equiv h(|m_q| + |m_{q'}|)(\delta f_{dc} - \delta f_{q,q'})$ between states $|q, L\rangle$ and $|q', R\rangle$ is proportional both to the net flux detuning from the location $\delta f_{q,q'}$ of the avoided crossing $A_{q,q'}$ and to the sum of the magnitudes of the energy-level slopes m_q and $m_{q'}$. Because the relative phase accumulated between the $|q, L\rangle$ and $|q', R\rangle$ components of the wavefunction over repeated Landau–Zener transitions is sensitive to $\Delta E_{q,q'}$, the slopes can be derived from the interference patterns which arise when δf_{dc} is varied. The N th node in the interference pattern, where a ‘node’ indicates a minimal change in the states’ populations, occurs when a relative phase of $2\pi N$ is accumulated between transitions²⁷ (Fig. 2b). For sinusoidal driving, the locations of the nodes (in δf_{dc}) follow the power law $s_{q,q'} N^{2/3}$, with a prefactor $s_{q,q'}$ related to the energy-level slopes by (see Supplementary Information)

$$|m_q| + |m_{q'}| = bv \sqrt{\frac{\alpha V}{s_{q,q'}^3}} \quad (4)$$

where $b = 3\pi/2\sqrt{2}$ and α is the frequency-dependent conversion factor between radio-frequency flux and source voltage; its value at $v = 0.16$ GHz, $\alpha = 0.082 \text{ m}\Phi_0 \text{ mV}_{\text{rms}}^{-1}$, is inferred from the slope of the left-hand edge of D1 (Fig. 2a). Figure 2b shows the $N^{2/3}$ power-law fits to the nodes of the vertical slices in D1 and D2, which are used to extract m_0 and m_1 (Fig. 2a, dashed vertical lines), where we take $|m_q| = |m_{q'}| \equiv m_q$ for $q = q'$ in our qubit. The slopes are obtained sequentially from the fitted values $s_{q,q'}$ in equation (4), starting with $2m_0 = 2.88 \text{ GHz m}\Phi_0^{-1}$ followed by $m_0 + m_1 = 2.534 \text{ GHz m}\Phi_0^{-1}$; their values are summarized in Table 1.

As an alternative way to analyse the data, we use the discrete two-dimensional Fourier transform (2DFT). To see the benefits of the 2DFT, we note that the amplitude spectroscopy plots in Figs 1 and 2 display structure on several scales. On the largest scale, the boundaries of the spectroscopy diamonds are readily identifiable, and on a smaller scale, the interiors of the diamonds are textured by fringes arising from the interference between successive Landau–Zener transitions at a single or multiple avoided crossings. On an even smaller scale, these fringes are composed of a series of horizontal multiphoton resonance lines. To extract information from these small-scale structures, it is helpful to apply a transformation that is able to invert length scales; the 2DFT does this.

In particular, as illustrated in Fig. 2c, the 2DFT allows us to determine the relation between the slopes m_0 and m_1 in a very clear and direct fashion (see Supplementary Information). The observed structure in the first two diamonds arises from the underlying ‘Bessel-function staircases’ of multiphoton resonances associated with transitions between the four lowest energy levels, where the n -photon absorption rate depends on driving amplitude through the square of the N th-order Bessel function^{22,27–30}. Using Fourier analysis, the apparently complicated mesh of overlapping Bessel functions is transformed to a pair of sinusoids $k_V = \pm \alpha g \sin(k_{\delta f_{dc}}/g)$, where $g = 2(|m_q| + |m_{q'}|)/v$, with periodicity related to the energy-level slopes⁴⁶. The sinusoid associated with $q = q' = 0$ arises from the transitions at $A_{0,0}$, whereas the second sinusoid, associated with $q = 0, q' = 1$ and $q = 1, q' = 0$, is degenerate and arises from the transitions at $A_{0,1}$ and $A_{1,0}$. Thus, the half-periods marked in Fig. 2c are $k_0 = 4\pi|m_0|/v$ and $k_1 = 2\pi(|m_0| + |m_1|)/v$. All four diamonds and their individual Fourier transforms are presented in Supplementary Figs 2–5.

Time-dependent amplitude spectroscopy

Valuable additional information about the energy-level spectrum and temporal coherence is gained by performing amplitude spectroscopy over short timescales (Fig. 3). Rather than the time-averaged

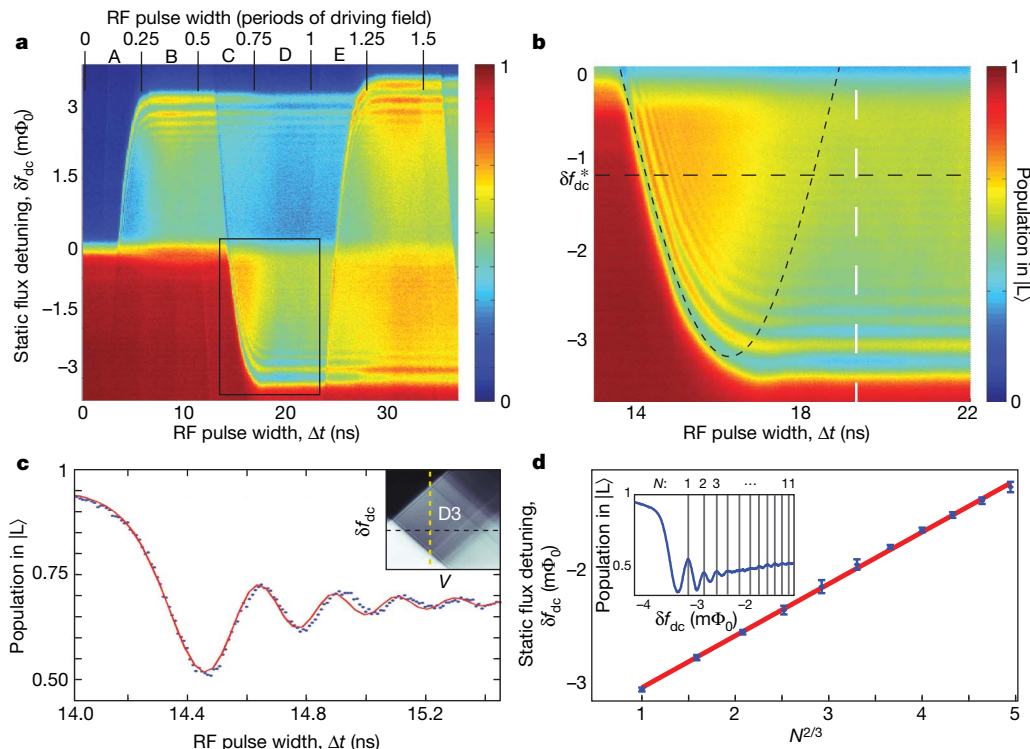


Figure 3 | Amplitude spectroscopy with short-pulse driving. **a**, Qubit response to a short radio-frequency pulse (equation (1)) of variable width Δt as a function of static flux detuning δf_{dc} , with $V = 181 \text{ mV}_{\text{rms}}$ and $\nu = 0.045 \text{ GHz}$. The scan is positioned on the left-hand side of D3 (inset of **c**; enlarged in Supplementary Fig. 7a), reaching all level crossings through $\Delta_{0,2}$ and $\Delta_{2,0}$. The top axis shows pulse width displayed in quarter-period regions A–E. **b**, Detail of the interference pattern in the outlined region of **a**. The black parabola marks the pulse widths at which the sinusoidal flux

excursion first exceeds, and then returns through, $\Delta_{0,2}$. **c**, Temporal oscillations along the horizontal line in **b** at $\delta f_{dc} = \delta f_{dc}^*$, fitted by a Landau–Zener model with damping (red line; see text and Supplementary Information). The inset shows D3 with the chosen V and δf_{dc} indicated by dashed lines. Grey scale is as in Fig. 4a. **d**, Interference-node positions versus $N^{2/3}$ along the vertical line in **b**, and best linear fit. The error bars indicate residual estimates from identifying node positions. The inset shows the interference pattern along the vertical line in **b**, and the node locations.

population discussed above, the time-dependent technique allows us to observe Larmor-type oscillations in population in the time domain and the real-time build-up of Stückelberg oscillations, even in systems with coherence times shorter than the driving period. In this measurement, we initialize the system to the ground state at a given detuning δf_{dc} , and then apply a harmonic driving pulse of a variable length Δt with fixed frequency ν and amplitude V , of the form given in equation (1). The phase of the sinusoid at the onset of each pulse is carefully adjusted to maintain the timing and directionality of the radio-frequency flux excursion through the energy levels. After the pulse ends abruptly at $t = \Delta t$ and $\delta f(t)$ returns to δf_{dc} , the qubit magnetization remains approximately frozen for times shorter than the inter-well relaxation time.

The main features of the time-dependent response are illustrated in Fig. 3, where parameters are tuned to investigate the $\Delta_{2,0}$ level crossing (Fig. 3c, inset). At positive flux detuning $\delta f_{dc} > 0$, the qubit is initialized to the ground state $|0, R\rangle$, whereas at $\delta f_{dc} < 0$, the ground state is $|0, L\rangle$ (Fig. 1c). Because in our qubit the splittings $\Delta_{0,0}$ and $\Delta_{1,0}$ ($\Delta_{0,1}$) are much smaller than $\Delta_{2,0}$ ($\Delta_{0,2}$), the change in qubit population per driving cycle is dominated by Landau–Zener transitions at the crossings $\Delta_{2,0}$ and $\Delta_{0,2}$. For positive flux detuning $\delta f_{dc} > 0$, the qubit is driven through $\delta f(t) < 0$, with significant population transfer first occurring in region A when $\Delta_{2,0}$ is reached (Fig. 3a). The onset of population transfer $|0, R\rangle \rightarrow |2, L\rangle$ is followed by brief, temporal Larmor-type oscillations between these states (see below). The population becomes stationary after the qubit returns through $\Delta_{2,0}$ in the second quarter-period (region B).

Because excited-state population in $|2, L\rangle$ tends to relax to the ground state $|0, L\rangle$, the next prominent population transfer, $|0, L\rangle \rightarrow |2, R\rangle$, occurs when the qubit is subsequently driven through

the avoided crossing $\Delta_{0,2}$ (positioned symmetrically to $\Delta_{2,0}$ in the energy-level diagram in Fig. 1c). This Landau–Zener transition, observed in the third quarter-period (region C), is again followed by intra-well relaxation from $|2, R\rangle$ to $|0, R\rangle$. The population then remains nearly constant (region D) until a third abrupt population transfer occurs during the first quarter of the second period (region E), which signals the qubit's return to $\Delta_{2,0}$ and the repetition of the driving cycle. The range in flux where the population transfer occurs during the first half-period is not as wide as it is for subsequent half-periods, because our pulse has slightly lower amplitude for times $\Delta t < 5 \text{ ns}$.

The observed response is not symmetric about $\delta f_{dc} = 0$. When starting at negative static bias $\delta f_{dc} = \delta f_{dc}^* < 0$, under harmonic driving (Fig. 1c, green sinusoid), the system is first drawn deeper into the ground state during the first half-period, without any level crossings. It is only during the second half-period that crossing $\Delta_{0,2}$ is reached and the first significant Landau–Zener transition occurs. The detailed time dependence of the population in this interval is shown in Fig. 3b. We can extract an approximate value of $\Delta_{0,2}$ by fitting the observed population change to equation (3), and obtain a refined estimate through the simulation described below.

The temporal oscillations, or ‘ringing’, displayed in Fig. 3b, c can be understood qualitatively in a pseudo-spin-1/2 picture, in which the qubit states are identified with up- and down-spin states relative to a fictitious z axis. The qubit undergoes Larmor-type precession about a tilted effective magnetic field which steadily increases in magnitude and rotates towards the z axis as the qubit leaves the avoided-crossing region. This picture is consistent with a temporal analysis of the canonical Landau–Zener problem, in which a linear ramp with velocity β sweeps the qubit through the avoided crossing.

In the perturbative (non-adiabatic) limit, this model yields the transition probability

$$P(t) = \frac{A_{0,2}^2}{4\hbar^2} \left| \int_{-\infty}^t e^{i\beta t'^2/2\hbar} dt' \right|^2 \quad (5)$$

(see Supplementary Information). The integral in equation (5) often arises in the context of optical diffraction, where it gives rise to Fresnel oscillations similar to the coherent oscillations observed in Fig. 3c.

Although equation (5), with $\beta = \zeta_i$ given in equation (2), captures the essential features of the data in Fig. 3c, to obtain a quantitative fit we must account for decoherence and the non-abrupt ending of the pulse, which adds a small Stückelberg-type interference²⁷ contribution (see Supplementary Information). We find good agreement between the data and a simulation of the Bloch dynamics of the two-level system near $A_{0,2}$, which includes longitudinal sinusoidal driving up to time $t = \Delta t$ followed by a rapid turn-off transient over approximately 2 ns, and a decoherence rate of $2\pi \times 0.65 \text{ ns}^{-1}$ (Fig. 3c). This large value is dominated by intra-well relaxation and phase jitter. The value of $A_{0,2}$ can be extracted as a fitting parameter and, in this regime, is largely insensitive to the details of the pulse transient and decoherence. Although the resulting coherence times are relatively short in comparison with the driving period, they are comparable to the typical Larmor frequency, set by the sweep rate, which allows us to observe coherent oscillations. Furthermore, as the qubit is swept back through the $A_{0,2}$ crossing, the interference that occurs at the second Landau–Zener transition mediates the conversion of temporal Larmor-type oscillations into Stückelberg steady-state oscillations.

As in the case of the stationary driving in Fig. 2, the energy-level slopes can be extracted from the Stückelberg fringes (Fig. 3b) using the $N^{2/3}$ power-law fitting (Fig. 3d) and equation (4). We infer m_2 and m_3 from the sums $m_0 + m_2 = 2.189 \text{ GHz m}\Phi_0^{-1}$ and $m_0 + m_3 = 1.929 \text{ GHz m}\Phi_0^{-1}$. We use the short-time amplitude spectroscopy procedure to obtain $A_{q,q'}$ for D2–D4 and slopes m_q for D3 and D4, as summarized in Table 1 ($A_{0,0}$ is obtained using the method of ref. 29).

Transverse modes

So far, we have focused only on the strongly coupled longitudinal modes. However, the lack of perfect symmetry allows us also to probe excited transverse modes within our driving scheme. The population transfer is relatively weak, indicating small deviations from an ideally symmetric double-well potential and longitudinal driving. Signatures of these states appear in D3 and D4 (see, for example, Fig. 1a; Fig. 4b, inset; and Supplementary Fig. 8a). The temporal response to a short radio-frequency pulse measured for an amplitude on the left-hand side of D3 (Fig. 4b, inset) at positive flux detuning is shown in Fig. 4a. The left-hand side of D3 marks the crossing $A_{02,00}$ of the states $|0, 0, L\rangle$ and $|0, 2, R\rangle$ during the first half-period, where some population is transferred from the right-hand well to the left-hand well, with the associated intra-well relaxation to $|0, 0, L\rangle$ (and we use the full level-crossing notation, explicitly indicating both longitudinal and transverse modes). During the second half-period, two weak population transfers are identified between the known positions of the longitudinal avoided crossings $A_{00,02}$ and $A_{00,03}$. This result is in agreement with simulations of the qubit Hamiltonian^{42,43}, which indicate that two transverse modes, $|1, 2, R\rangle$ and $|2, 2, R\rangle$, exist in this region, as illustrated in Fig. 4b. Although we can identify their locations, the values of $A_{00,12}$ and $A_{00,22}$ are not conclusively determined from this measurement, because the population change is small compared with that of the adjacent longitudinal crossings $A_{00,02}$ and $A_{00,03}$.

Concluding remarks

The amplitude spectroscopy demonstrated here is complementary to conventional frequency spectroscopy; it is generally applicable to systems with traversable avoided crossings, including both artificial

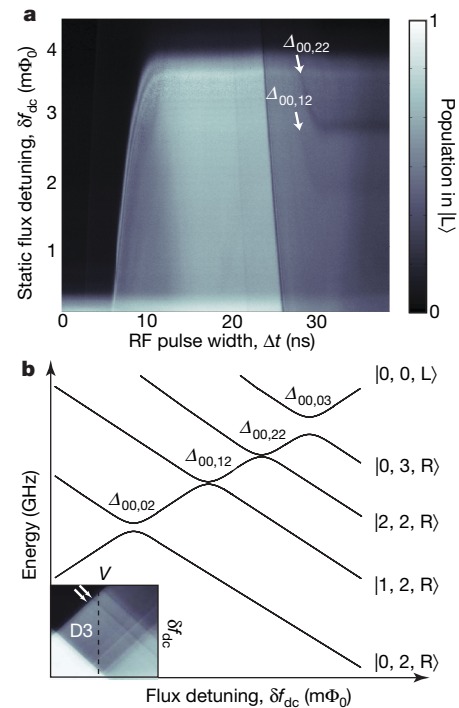


Figure 4 | Identification of transverse qubit states. **a**, Qubit response to a short radio-frequency pulse of variable length Δt as a function of static flux detuning δf_{dc} , with $V = 179 \text{ mV}_{\text{rms}}$ and $\nu = 0.025 \text{ GHz}$. The scan is positioned on the left-hand side of D3 (inset of **b**; enlarged in Supplementary Fig. 7b), where the crossing $A_{03,00}$ (but not $A_{00,03}$) is reached. The signatures of two crossings with transverse states, $A_{00,12}$ and $A_{00,22}$, are indicated with arrows here and in the inset of **b**. **b**, Energy-level diagram showing the locations of the transverse states. The inset shows D3 with the chosen V indicated by a dashed line. Grey scale is as in **a**.

and natural atomic systems. Owing to the sensitivity of interference conditions and transition probabilities to system parameters, it is a useful tool in the study and manipulation of quantum systems, and we can envision it opening new pathways for quantum control⁴⁷.

The amplitude spectroscopy technique can be utilized to study dissipative environments⁴⁸ by determining Landau–Zener probabilities at different sweep rates. It can also be extended to anharmonic excitation; for example, arbitrary-waveform excursions through the energy levels and targeted harmonic excitations can be used to achieve desired transitions. This type of hybrid driving was demonstrated very recently in caesium atoms⁴⁹ and rubidium atoms⁵⁰ about Feshbach resonances, which are systems containing weakly coupled levels that are challenging to address within the standard frequency-based approach.

METHODS SUMMARY

Qubit readout. The qubit states are read out using a d.c. SQUID, a sensitive magnetometer that distinguishes the flux generated by the qubit persistent currents I_q . The readout is performed by driving the SQUID with a current pulse I_{SQ} comprising a 20-ns ‘sample’ current I_s followed by a 20- μs ‘hold’ current (Fig. 1b). The SQUID will switch to its normal-state voltage if $I_s > I_{sw,L}$ or $I_s > I_{sw,R}$ when the qubit is in the respective state $|L\rangle$ or $|R\rangle$. By sweeping the SQUID sample current I_s and qubit flux detuning δf_{dc} while monitoring the presence of a SQUID voltage V_{SQ} over many trials, we generate a cumulative switching-current distribution function. From this distribution, we extract a best-estimator line in the space of I_s and δf_{dc} that allows us to characterize the population of state $|L\rangle$ for a given flux detuning.

Experiment implementation. The experiments were performed in a dilution refrigerator at a temperature of approximately 20 mK. The device was magnetically shielded using four Cryoperm 10 cylinders and a superconducting enclosure. All electrical leads were attenuated and/or filtered to minimize noise. The electrical temperature of the device in the absence of microwave cooling was approximately 40 mK. After the microwave cooling pulse had been applied

(Fig. 1b), the effective temperature of the qubit was less than 3 mK. Microwave cooling enabled the data to be obtained at a repetition rate of 10 kHz, which is generally faster than the intrinsic equilibration rate due to inter-well relaxation. For all experiments, the static flux detuning was swept in $60\text{-}\mu\Phi_0$ steps, and the radio-frequency pulse amplitude was scanned in $0.5\text{-mV}_{\text{rms}}$ steps (at the source). The pulse width was scanned in steps of 0.005 ns to 0.1 ns, and each data point comprised an average of 500 to 30,000 trials, depending on the desired resolution.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 May; accepted 11 July 2008.

- Schawlow, A. L. Spectroscopy in a new light. *Rev. Mod. Phys.* **54**, 697–707 (1982).
- Thompson, R. C. High resolution laser spectroscopy of atomic systems. *Rep. Prog. Phys.* **48**, 531–578 (1985).
- Friedman, J. R., Patel, V., Chen, W., Tolpygo, S. K. & Lukens, J. E. Quantum superposition of distinct macroscopic states. *Nature* **406**, 43–46 (2000).
- van der Wal, C. H. *et al.* Quantum superposition of macroscopic persistent-current states. *Science* **290**, 773–777 (2000).
- Berkley, A. J. *et al.* Entangled macroscopic quantum states in two superconducting qubits. *Science* **300**, 1548–1550 (2003).
- Izmalkov, A. *et al.* Evidence for entangled states of two coupled flux qubits. *Phys. Rev. Lett.* **93**, 037003 (2004).
- van der Wiel, W. G. *et al.* Electron transport through double quantum dots. *Rev. Mod. Phys.* **75**, 1–22 (2003).
- Clarke, J., Cleland, A. N., Devoret, M. H., Esteve, D. & Martinis, J. H. Quantum mechanics of a macroscopic variable: the phase difference of a Josephson junction. *Science* **239**, 992–997 (1988).
- Hanson, R., Kouwenhoven, L. P., Petta, J. R., Tarucha, S. & Vandersypen, L. M. K. Spins in few-electron quantum dots. *Rev. Mod. Phys.* **79**, 1217–1265 (2007).
- Nakamura, Y., Pashkin, Y. A. & Tsai, J. S. Coherent control of macroscopic quantum states in a single-Cooper-pair box. *Nature* **398**, 786–788 (1999).
- Vion, D. *et al.* Manipulating the quantum state of an electrical circuit. *Science* **296**, 886–889 (2002).
- Yu, Y., Han, S., Chu, X., Chu, S.-I. & Wang, Z. Coherent temporal oscillations of macroscopic quantum states in a Josephson junction. *Science* **296**, 889–892 (2002).
- Martinis, J. M., Nam, S. & Aumentado, J. Rabi oscillations in a large Josephson-junction qubit. *Phys. Rev. Lett.* **89**, 117901 (2002).
- Chiorescu, I., Nakamura, Y., Harmans, C. J. P. M. & Mooij, J. E. Coherent quantum dynamics of a superconducting flux qubit. *Science* **299**, 1869–1871 (2003).
- Collin, R. E. *Foundations for Microwave Engineering* (Wiley-IEEE, 2001).
- Friedman, J. R. *et al.* Macroscopic measurement of resonant magnetization tunnelling in high-spin molecules. *Phys. Rev. Lett.* **76**, 3830–3833 (1996).
- Thomas, L. *et al.* Macroscopic quantum tunnelling of magnetization in a single crystal of nanomagnets. *Nature* **383**, 145–147 (1996).
- Wernsdorfer, W. & Sessoli, R. Quantum phase interference and parity effects in magnetic molecular clusters. *Science* **284**, 133–135 (1999).
- Cohen-Tannoudji, C., Dupont-Roc, J. & Grynberg, G. *Atom-Photon Interactions: Basic Processes and Applications* Ch. 6 (Wiley, 1992).
- Nakamura, H. *Nonadiabatic Transition* Ch. 12 (World Scientific, 2001).
- Stückelberg, E. C. G. Theory of inelastic collisions between atoms. *Helv. Phys. Acta* **5**, 369–422 (1932).
- Nakamura, Y., Pashkin, Y. A. & Tsai, J. S. Rabi oscillations in a large Josephson-junction charge two-level system. *Phys. Rev. Lett.* **87**, 246601 (2001).
- Claudon, J., Balestro, F., Hekking, J. W. J. & Buisson, O. Coherent oscillations in a superconducting multilevel quantum system. *Phys. Rev. Lett.* **93**, 187003 (2004).
- Plourde, B. L. T. *et al.* Flux qubits and readout device with two independent flux lines. *Phys. Rev. B* **72**, 060506(R) (2005).
- Saito, S. *et al.* Parametric control of a superconducting flux qubit. *Phys. Rev. Lett.* **96**, 107001 (2006).
- Izmalkov, A. *et al.* Observation of macroscopic Landau-Zener transitions in a superconducting device. *Eur. Phys. Rev. Lett.* **65**, 844–849 (2004).
- Oliver, W. D. *et al.* Mach-Zehnder interferometry in a strongly driven superconducting qubit. *Science* **310**, 1653–1657 (2003).
- Sillanpää, M., Lehtinen, T., Paila, A., Makhlin, Yu & Hakonen, P. Continuous-time monitoring of Landau-Zener interference in a Cooper-pair box. *Phys. Rev. Lett.* **96**, 187002 (2006).
- Berns, D. M. *et al.* Coherent quasiclassical dynamics of a persistent current qubit. *Phys. Rev. Lett.* **97**, 150502 (2006).
- Wilson, C. M. *et al.* Coherence times of dressed states of a superconducting qubit under extreme driving. *Phys. Rev. Lett.* **98**, 257003 (2007).
- Valenzuela, S. O. *et al.* Microwave-induced cooling of a superconducting qubit. *Science* **314**, 1589–1592 (2006).
- Niskanen, A. O., Nakamura, Y. & Pekola, J. P. Information entropic superconducting microcooler. *Phys. Rev. B* **76**, 174523 (2007).
- You, J. Q., Liu, Y. & Nori, F. Simultaneous cooling of an artificial atom and its neighboring quantum system. *Phys. Rev. Lett.* **100**, 047001 (2008).
- Chiorescu, I. *et al.* Coherent dynamics of a flux qubit coupled to a harmonic oscillator. *Nature* **431**, 159–162 (2004).
- Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
- Johansson, J. *et al.* Vacuum Rabi oscillations in a macroscopic superconducting qubit LC oscillator system. *Phys. Rev. Lett.* **96**, 127006 (2006).
- Sillanpää, M. A., Park, J. I. & Simmonds, R. W. Coherent quantum state storage and transfer between two phase qubits via a resonant cavity. *Nature* **449**, 438–442 (2007).
- Majer, J. *et al.* Coupling superconducting qubits via a cavity bus. *Nature* **449**, 443–447 (2007).
- Siddiqi, I. *et al.* RF-driven Josephson bifurcation amplifier for quantum measurement. *Phys. Rev. Lett.* **93**, 207002 (2004).
- Katz, N. *et al.* Coherent state evolution in a superconducting qubit from partial-collapse measurement. *Science* **312**, 1498–1500 (2006).
- Lupaşcu, A. *et al.* Quantum non-demolition measurement of a superconducting two-level system. *Nature Phys.* **3**, 119–125 (2007).
- Mooij, J. E. *et al.* Josephson persistent-current qubit. *Science* **285**, 1036–1039 (1999).
- Orlando, T. P. *et al.* Superconducting persistent-current qubit. *Phys. Rev. B* **60**, 15398–15413 (1999).
- Mark, M. *et al.* “Stueckelberg interferometry” with ultracold molecules. *Phys. Rev. Lett.* **99**, 113201 (2007).
- Astafiev, O. *et al.* Single artificial-atom lasing. *Nature* **449**, 588–590 (2007).
- Rudner, M. S. *et al.* Quantum phase tomography of a strongly driven qubit. Preprint at (<http://arxiv.org/abs/0805.1555>) (2008).
- Ashab, S., Johansson, J. R., Zagorin, A. M. & Nori, F. Two-level systems driven by large-amplitude fields. *Phys. Rev. A* **75**, 063414 (2007).
- Wubs, M. *et al.* Gauging a quantum heat bath with dissipative Landau-Zener transitions. *Phys. Rev. Lett.* **97**, 200404 (2006).
- Mark, M. *et al.* Spectroscopy of ultracold trapped caesium Feshbach molecules. *Phys. Rev. A* **76**, 042514 (2007).
- Lang, F. *et al.* Cruising through molecular bound-state manifolds with radiofrequency. *Nature Phys.* **4**, 223–226 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Shytov, J. Bylander, B. Turek, A. J. Kerman and J. Sage for discussions and D. Baker, V. Bolkhovsky, G. Fitch, E. Macedo, P. Murphy, K. Parrillo, R. Slattery and T. Weir at Lincoln Laboratory, MIT, for technical assistance. This work was supported by the Air Force Office of Scientific Research and the Laboratory for Physical Sciences (F49620-01-1-0457) under the Defense University Research Initiative in Nanotechnology programme, and by the US government. The work at Lincoln Laboratory was sponsored by the US Department of Defence under Air Force Contract No. FA8721-05-C-0002.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.D.O. (oliver@ll.mit.edu).

METHODS

Device fabrication and parameters. The device was fabricated at Lincoln Laboratory, MIT. It has a critical current density of $J_c \approx 160 \text{ A cm}^{-2}$, and the characteristic Josephson and charging energies are $E_J \approx 2\pi \hbar \times 300 \text{ GHz}$ and $E_C \approx 2\pi \hbar \times 0.65 \text{ GHz}$, respectively. The ratio of the qubit Josephson junction areas is approximately 0.84. The qubit loop area is $16 \times 16 \mu\text{m}^2$, and its self-inductance is $L_q \approx 30 \text{ pH}$. The SQUID Josephson junctions each have critical current $I_{c0} \approx 2 \mu\text{A}$. The SQUID loop area is $20 \times 20 \mu\text{m}^2$, and its self-inductance is $L_{SQ} \approx 30 \text{ pH}$. The SQUID junctions were shunted with two 1-pF on-chip capacitors. The mutual coupling between the qubit and the SQUID is $M \approx 25 \text{ pH}$.

Potential energy of the persistent-current qubit. The potential energy of the persistent-current qubit is a two-dimensional anisotropic periodic potential with double-well structures at each lattice site^{42,43}. It was designed to have negligible inter-lattice-site tunnelling, so the potential energy can be visualized as a single double-well, as seen in Supplementary Fig. 1a. We parameterize the potential energy U using the phase variables $\varphi_m = (\varphi_1 - \varphi_2)/2$ and $\varphi_p = (\varphi_1 + \varphi_2)/2$, where φ_1 and φ_2 are the phases across the two largest of the three junctions^{42,43} (Fig. 1b). It is convenient to plot U/E_J , where $E_J = (I_c/2\pi) \Phi_0$ and I_c is the critical current of the larger junctions. For an ideal double-well potential, the qubit is driven longitudinally, thereby conserving the parity of the transverse modes, and two-dimensional potential can be reduced to a one-dimensional double-well along the φ_m direction, as seen in Supplementary Fig. 1b. This is the longitudinal direction in which the qubit circulating current varies through the phase φ_m (refs 42, 43).

ARTICLES

Widespread changes in protein synthesis induced by microRNAs

Matthias Selbach¹, Björn Schwanhäusser^{1*}, Nadine Thierfelder^{1*}, Zhuo Fang¹, Raya Khanin² & Nikolaus Rajewsky¹

Animal microRNAs (miRNAs) regulate gene expression by inhibiting translation and/or by inducing degradation of target messenger RNAs. It is unknown how much translational control is exerted by miRNAs on a genome-wide scale. We used a new proteomic approach to measure changes in synthesis of several thousand proteins in response to miRNA transfection or endogenous miRNA knockdown. In parallel, we quantified mRNA levels using microarrays. Here we show that a single miRNA can repress the production of hundreds of proteins, but that this repression is typically relatively mild. A number of known features of the miRNA-binding site such as the seed sequence also govern repression of human protein synthesis, and we report additional target sequence characteristics. We demonstrate that, in addition to downregulating mRNA levels, miRNAs also directly repress translation of hundreds of genes. Finally, our data suggest that a miRNA can, by direct or indirect effects, tune protein synthesis from thousands of genes.

MicroRNAs are key *trans*-acting factors that post-transcriptionally regulate metazoan gene expression, and identifying miRNA targets as well as the effect that miRNAs exert on them is a fundamental question for understanding life, health and disease^{1–5}. The first identified miRNA targets in *Caenorhabditis elegans* were found to be translationally repressed whereas target mRNA levels were only mildly downregulated. Subsequently, similar cases were reported in mammalian systems^{6,7}. Reporter constructs provided experimental evidence that miRNAs can directly repress translation initiation^{8–10}. Furthermore, it has been shown that different mechanisms exist by which miRNAs repress protein synthesis or induce mRNA degradation^{6,11}. Overexpressing a miRNA in human cell lines causes mostly mild (less than twofold) downregulation of hundreds of mRNAs, of which many are direct targets¹². Nonetheless, these results do not reveal how much control miRNAs exert on protein synthesis. Because protein synthesis is one of the most important quantities for the phenotype, a fundamental question about gene regulation has therefore remained unanswered.

Identifying miRNA targets has been the subject of a steeply growing number of computational^{13–16} and experimental^{17–20} approaches. Although certain features of the miRNA-binding site such as seed sites (Watson–Crick consecutive base pairing between mRNAs and the miRNA at position 2–7 counted from its 5' end) located in the 3' untranslated regions (3' UTRs) of mRNAs are important, it is unknown how relevant they are for changes in protein production. Several rules regarding the architecture of miRNA-binding sites have been proposed to explain differences in their efficacy in mRNA degradation versus translational repression^{6,21}. However, these rules were based on a few target sites that were studied mostly in reporter assays with non-endogenous proteins. Another study about the effects of miRNA on the proteome was limited by the small number (12) of detected downregulated proteins²². Furthermore, different proteins have different turnover times. For example, if a miRNA completely shuts off protein production, steady-state levels of high-turnover proteins will change rapidly whereas stable proteins will be affected later. Therefore, changes in protein concentrations as measured by standard techniques cannot quantify changes in protein synthesis if protein

levels are not stationary. In fundamental biological processes such as differentiation, the expression of miRNAs is strongly induced (or switched off) in a relatively small time window²³. Thus, to assess endogenous regulation of mRNA translation by miRNAs, a technique is needed to measure directly genome-wide changes in protein synthesis shortly after changes in miRNA expression.

pSILAC measures changes in protein production

To overcome these problems, we devised a new variant of SILAC (stable isotope labelling with amino acids in cell culture). In SILAC, proteins are metabolically labelled by cultivating cells in growth medium containing heavy isotope versions of essential amino acids^{24,25}. Mass spectrometry can distinguish peptides derived from SILAC-labelled proteins. The ratio of peptide peak intensities reflects differences in corresponding protein abundance. We reasoned that by pulse-labelling with two different heavy stable isotope labels we could measure changes in protein production between two samples. In our pulsed SILAC (pSILAC) method, cells in the two samples are pulse-labelled with two different heavy versions of amino acids. During labelling, all newly synthesized proteins will be 'heavy' or 'medium-heavy' (Fig. 1a). Pre-existing proteins present before labelling remain in the light form and are ignored. Only intensity differences between newly synthesized proteins (medium-heavy and heavy) are considered. Hence, pSILAC quantifies differences in protein production between both samples integrated over the measurement time after the pulse²⁶. This is fundamentally different from pulse-labelling with a single label to determine protein turnover or transport^{27–29}. We combined pSILAC with state-of-the-art mass-spectrometry-based proteomics^{30–32} to measure changes in production of ~5,000 proteins altogether.

We performed transfections to individually overexpress five human miRNAs in HeLa cells. These miRNAs are tissue-specific and virtually absent in HeLa cells (miR-1, miR-155) or expressed in many tissues (miR-16, miR-30a, let-7b) including HeLa cells³³. At least 90% of all cells could be efficiently transfected (Supplementary Fig. 1), and miRNAs were overexpressed for at least 32 h post-transfection (not shown). Changes in protein production

¹Max Delbrück Center for Molecular Medicine, Robert-Rössle Str. 10, D-13125 Berlin, Germany. ²Department of Statistics, 15 University Gardens, University of Glasgow, Glasgow, G12 8QQ, UK.

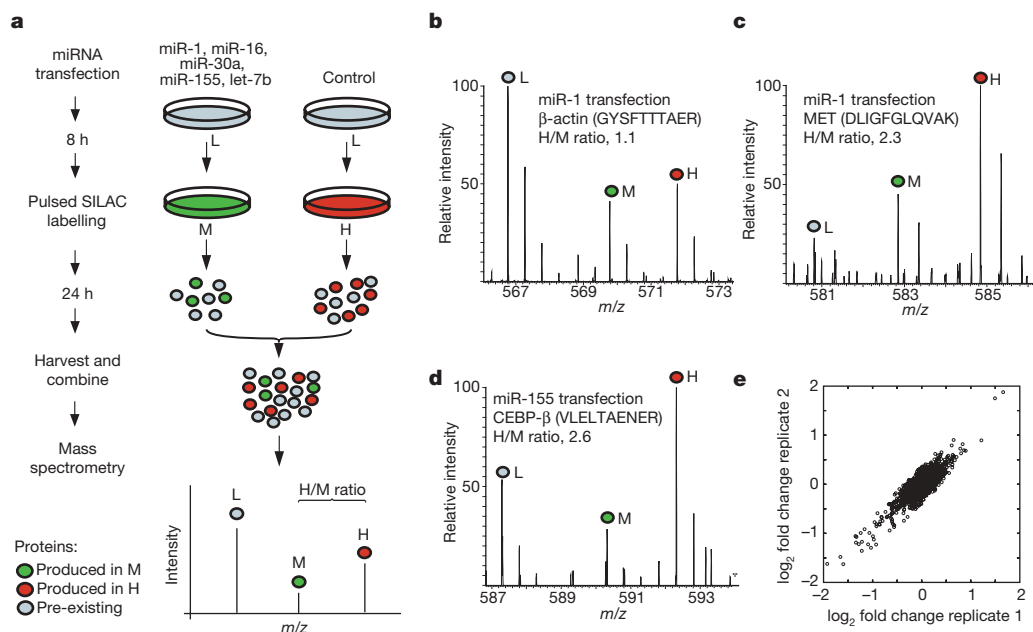


Figure 1 | Global analysis of changes in protein production induced by microRNAs. **a**, HeLa cells cultivated in normal light (L) medium were either transfected with a miRNA or mock transfected. After 8 h, transfected and control cells were pulse-labelled by transferring them to culture medium containing medium-heavy (M) or heavy (H) isotope-labelled amino acids, respectively (pSILAC). All newly synthesized proteins will appear in the H or M form. Samples were combined after 24 h and analysed by mass

spectrometry. Intensity peak ratios between heavy and medium-heavy peptides (H/M ratio) reflect changes in protein production. RNA from the same samples was analysed by microarrays. **b–d**, Exemplary peptide mass spectra (sequences are in parentheses). The production of most proteins is unaltered, as shown for a β-actin peptide. In contrast, synthesis of MET and CEBPβ is reduced by miR-1 or miR-155 overexpression. **e**, Reproducibility of pSILAC (biological replicate, see Supplementary Methods).

were measured by pulse-labelling at 8 h post-transfection over a time period of 24 h. Representative mass spectra are shown in Fig. 1b–d. In total, we identified 4,961 proteins in HeLa cells with high confidence (false discovery rate <1%, see Supplementary Methods). Although mass spectrometry is biased to detect more highly expressed genes,

this bias was mild and did not affect the detection range (Supplementary Fig. 2). We validated 16 out of 16 selected pSILAC measurements by western blotting (Supplementary Fig. 3). Analysis of biological replicates showed high correlation (Pearson correlation coefficient ~0.9) over the entire dynamic range (Fig. 1e).

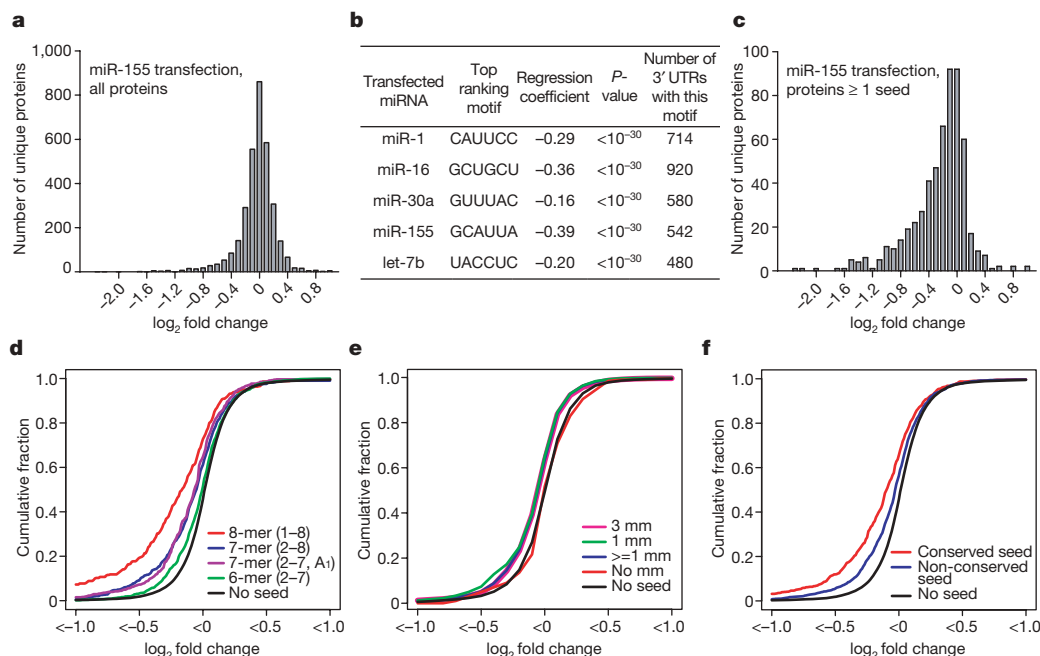


Figure 2 | miRNAs downregulate protein synthesis of hundreds of genes. **a**, Histogram of changes in production of 3,299 proteins in HeLa cells after miR-155 overexpression. **b**, An unbiased search for 3' UTR motifs that correlate with pSILAC fold changes yielded precisely the miRNA seed sequences. **c**, Proteins with miR-155 seeds tend to be downregulated by miR-155 overexpression. **d**, Cumulative distributions of different seed classes (matches to positions 1–8 (8-mer), 2–8 (7-mers), 2–7 with adenosine in

position 1 (2–7, A₁) and 2–7 (6-mer)). **e**, Mismatches (mm) between positions 9 and 11 of the miRNA and target mRNAs with a seed correlate with downregulation. Protein synthesis from mRNAs with perfect complementarity at positions 9–11 (red) and synthesis from mRNAs without seeds (black) is indistinguishable. **f**, Conserved seeds mediate more downregulation than non-conserved seeds. Results are shown for pooled data based on messages with one seed only (**d–f**).

mRNA sequence features of repressed proteins

Perhaps surprisingly, pSILAC revealed that miRNA overexpression had, overall, mild effects on the synthesis of most of the 3,000–3,500 proteins quantified in each transfection (shown for miR-155 in Fig. 2a). Because miRNAs are thought to target mRNAs primarily by binding *cis*-regulatory sites in 3' UTRs, we used a linear-regression-based analysis³⁴ to identify 3' UTR sequence motifs that correlated best with changes in protein production. This method performs an unbiased screen for all nucleotide motifs of one to six nucleotides in length. For each miRNA, the most significant motif of all possible 5,460 motifs was precisely the seed of the respective miRNA (Fig. 2b), and correlated with downregulation. The same motif search in 5' UTRs had no significant results. Searching coding sequences yielded the seed in only two experiments (let-7b, miR-16), and further analyses showed that 3' UTRs exert the strongest effect (Supplementary Fig. 4). Taking miR-155 as an example, the seed enrichment in downregulated proteins is illustrated by the histogram of fold changes for proteins that contain at least one seed in their mRNA 3' UTRs (Fig. 2c). Thus, proteins with reduced synthesis are enriched in direct miRNA targets, and a primary motif to mediate this reduction is the 3' UTR seed. Certain characteristics such as seed-flanking nucleotides have been reported to affect the degree of mRNA degradation by miRNAs^{35,36}, and we show that these effects are also involved in repressing protein production (Fig. 2d).

When small interfering RNAs (siRNAs) are perfectly complementary to their targets, mRNA cleavage occurs between nucleotides 10 and 11 opposite the siRNA guide strand; in contrast, mismatches in this region strongly reduce cleavage^{37–39}. A small-scale study with reporter constructs suggested that siRNA–mRNA pairs with mismatches between nucleotides 9–11 of the siRNA are mainly repressed at the protein level with little effect on the transcript²¹. We found that only seed-containing mRNAs with at least one mismatch were, overall, repressed at the protein level (Fig. 2e). In contrast, protein production from seed-containing mRNAs with perfect base pairing from nucleotides 9 to 11 and mRNAs lacking seeds was indistinguishable. Hence, although mismatches are deleterious to siRNA-mediated cleavage of mRNAs, they correlate with increased repression of protein production by miRNAs. We also found that, on average,

repression is more pronounced for conserved than for non-conserved seed sites (Fig. 2f), indicating that our experiments reflect biological relevance and that there are determinants in addition to the seed that mediate efficient downregulation of protein synthesis.

We next quantified how many of the downregulated proteins can be explained by the seed. We recorded how many proteins with at least one 3' UTR seed site were downregulated by at least *c*-fold as a function of *c* (Fig. 3a). For example, the production of more than 300 proteins with seeds was downregulated by at least 30% (\log_2 -fold change < -0.5). These proteins amounted to roughly 60–70% of all measured proteins downregulated by at least this much (Fig. 3b). Because the background seed frequency is 10–30% (Fig. 3b, dashed horizontal lines), we can explain up to 60% of the ~300 proteins by the presence of seeds. It remains an open question how many proteins without a seed are direct targets. Nevertheless, pSILAC clearly generates lists of proteins enriched in direct targets. We independently validated the 3' UTR-dependence of protein production by dual luciferase reporters for eight 3' UTRs with a seed for either miR-1 or let-7b (see Supplementary Methods). The correlation with the corresponding pSILAC data was high (Fig. 3c).

pSILAC data and target predictions

Having shown that pSILAC data are enriched in direct miRNA targets, we tested how miRNA target predictions correlate with our data. We calculated the fraction of predicted mRNA targets for which protein production was downregulated by at least *c*-fold. The results were consistent for all values of *c* and all miRNAs individually (data not shown). For example, roughly 27% of all 24,238 mRNAs present in the pSILAC data were downregulated more strongly than $-0.1 \log_2$ -fold change (Fig. 3d and Supplementary Table 1). A completely random selection would therefore have 27% overlap with pSILAC data. This background accuracy was exceeded by all methods except one based on 5' UTRs. Simply considering seed sites boosts the accuracy to 44%. This accuracy was only topped by three methods that use evolutionary conservation of seed sites as an additional filter. Almost all other methods, in part based on site-accessibility evaluation, made fewer predictions with less accuracy.

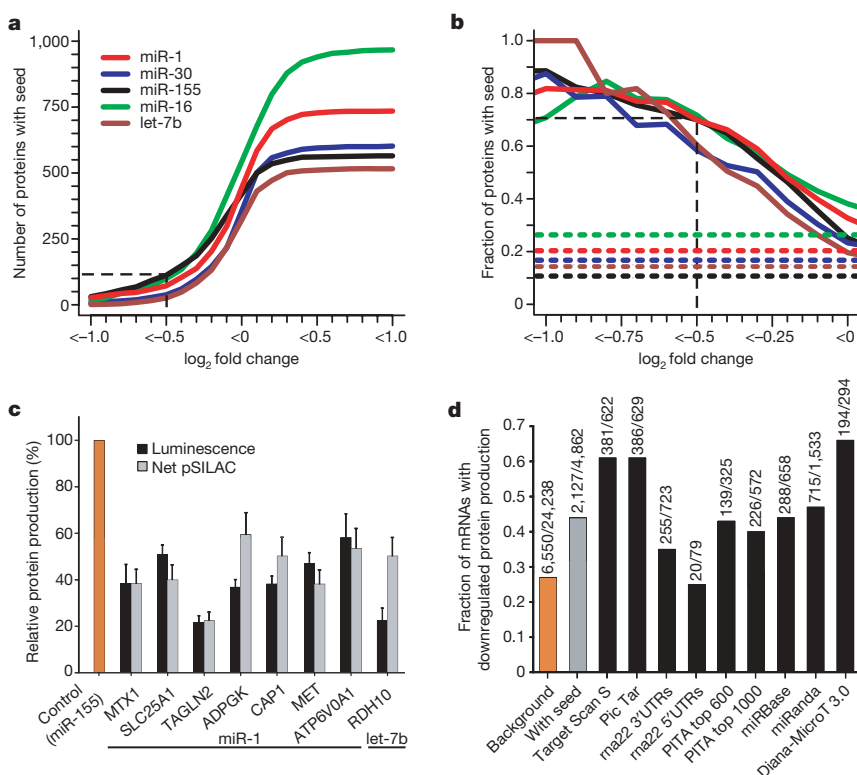


Figure 3 | The miRNA seed explains a large fraction of downregulated protein synthesis.

a, Cumulative number of proteins with seeds as a function of changes in their production. For a given cutoff, this indicates the number of downregulated seed-containing proteins (shown for $-0.5 \log_2$ -fold change). **b**, Fraction of proteins with a seed as a function of repression. Background seed frequency of unchanged proteins (absolute \log_2 -fold change < 0.1) ranges from 10–30% (dashed lines). **c**, Dual luciferase reporter assays for 3' UTR-mediated regulation by miRNAs (\pm s.d., $n = 3$). 'Net pSILAC' refers to the difference of pSILAC \log_2 -fold changes for the miRNA and the control (error bars show 95% confidence interval, see Supplementary Methods). **d**, The fraction of computationally predicted target mRNAs with reduced protein production (\log_2 -fold change < -0.1) is calculated for all five miRNA data sets pooled.

Translational repression by miRNAs

pSILAC measures changes in the amount of newly synthesized proteins between two samples. This depends on changes of mRNA levels and, in addition, on translational regulation. To discern these two mechanisms, in all pSILAC experiments we measured the mRNA fold changes between the miRNA-transfected sample and the control by Affymetrix microarrays at the beginning of the pulse labelling ($t_1 = 8$ h) and at the end ($t_2 = 32$ h). A total of 69 quantitative polymerase chain reactions with reverse transcription (qRT-PCRs) demonstrated that our microarray data have little compression or other distortion effects in the range where most mRNA fold changes were observed (Supplementary Fig. 5).

For miR-1 as an example, we present the relationship between miRNA-induced fold changes in protein production (pSILAC) and mRNA fold changes (Fig. 4a, b) separately for t_1 and t_2 . Very few genes had fold changes of unequal sign and reasonable magnitude (≥ 1.3 -fold). The correlation between mRNA fold changes and pSILAC fold changes became better at t_2 . In particular, many genes with downregulated protein production but little mRNA fold changes at t_1 shifted towards greater mRNA fold changes at t_2 . Similar overall effects could be seen for the other miRNAs. Nevertheless, the considerable scatter indicates substantial and widespread post-transcriptional regulation of gene expression.

The distribution of fold changes measured by microarray and pSILAC was similar (Fig. 4c, histograms). However, the average number (s) of seeds per gene was higher for more highly downregulated genes. Seed enrichment was not observed for upregulated genes, indicating that the recently reported miRNA-mediated activation of gene expression did not occur under our experimental conditions⁴⁰. For downregulated genes, log-fold changes were linearly correlated with s . Thus, if a target has two seeds, the repressive effect is multiplicative, as has been observed in small-scale studies^{15,41}. pSILAC data also support earlier findings³⁶ that synergistic effects are higher for two nearby seeds (< 40 nucleotides) compared to larger spacings (> 40 nucleotides; P -value 0.003, one-sided Wilcoxon test). Intriguingly, the slope of s in Fig. 4c is steeper for pSILAC fold changes, suggesting that the multiplicity of a miRNA-binding site

in the same 3' UTR exerts a stronger direct effect on protein production than on mRNA levels. To assess miRNA-mediated changes in translation rates for each gene, we subtracted the \log_2 mRNA from the \log_2 pSILAC fold changes, and plotted s as a function of these differences (Fig. 4d). The linear decay of s towards the regime of equal fold changes indicates that, in addition to mediating mRNA down-regulation¹², the seed also mediates direct repression of translation rates for hundreds of genes.

Endogenous miRNA knockdown

It could be argued that the overexpression of miRNAs can lead to largely non-physiological effects. We therefore used a locked nucleic acid (LNA) approach^{42,43} to knockdown let-7b in HeLa cells (Fig. 5a), and measured changes in protein production and mRNA levels as before. Luciferase reporter experiments demonstrated that our knockdown functionally derepressed a known let-7 target⁴⁴ mediated by seed sites (Supplementary Fig. 6). As in the overexpression experiments, an unbiased search for 3' UTR motifs identified the let-7b seed as the best match. Coding sequences and 5' UTRs did not yield significant results. Further analyses showed that all effects for seed-mediated targets that we report for the overexpression experiments hold true for the let-7b knockdown after flipping the sign of pSILAC and microarray fold changes, including correlation of target-finding algorithms with pSILAC data (Supplementary Fig. 7). Together, these data suggest that the miRNA overexpression experiments are also physiologically relevant.

let-7b tunes production of thousands of proteins

When we compared the cellular response to let-7b overexpression and knockdown we observed a marked anti-correlation, not only for seed-mediated let-7b targets but also for most of the $\sim 2,700$ proteins quantified in both experiments (that is, for both direct and indirect effects; Fig. 5b). For example, when considering all ~ 130 proteins with a fold change of at least 15% in both the overexpression and knockdown experiments, most were up in one of the experiments but down in the other, irrespective of seeds (Fig. 5c). In contrast, almost all proteins with seeds were down in the overexpression experiment

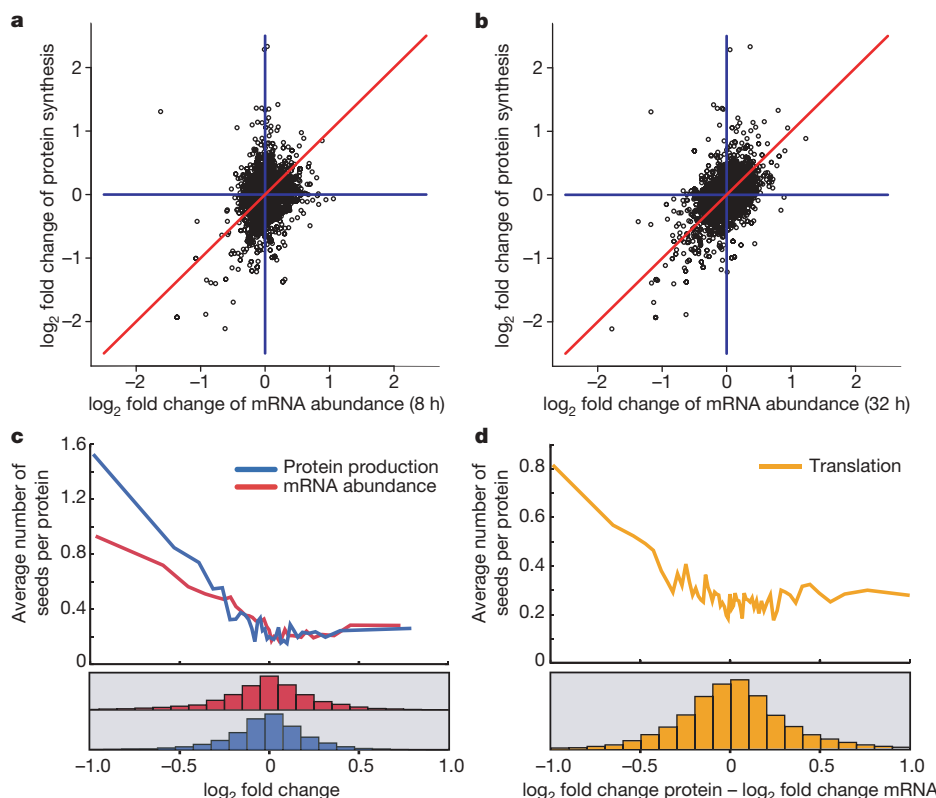


Figure 4 | miRNAs inhibit translation on a genome-wide scale. **a**, Changes in protein production between 8 h and 32 h after miR-1 transfection with mRNA fold changes at 8 h reveal poor overall correlation. **b**, mRNA levels at 32 h correlate remarkably well with changes in protein synthesis. **c**, Overall fold changes of mRNA and protein synthesis are similar (histograms). Reduced protein production and mRNA levels correlate with seed frequency (curves represent proteins ranked by fold change and grouped into bins of 250). **d**, Translational repression by miRNAs is revealed by subtracting mRNA log changes from log changes in protein production. Increased seed frequency, averaged as in **c**, correlates with translational repression. Results are shown for pooled data (**c**, **d**) after discarding genes with mRNA and pSILAC changes of unequal sign.

and up in the knockdown. When averaging the data, we found a linear response of the entire proteome to miRNA misexpression with a slope of -0.3 (Fig. 5b, inset), demonstrating that, on average, let-7b overexpression induced roughly threefold higher \log_2 -fold changes than let-7b knockdown. Together, these data indicate that upregulation and downregulation of stationary let-7b levels has largely complementary effects on the proteome; that is, let-7b levels can tune protein production from thousands of genes.

Discussion

Here we have measured for the first time changes in cellular protein synthesis in response to miRNA induction or knockdown on a proteome-wide scale. Our results show that a single miRNA can directly downregulate production of hundreds of proteins. In addition to the known effect on global mRNA levels¹², our data strongly indicate that miRNAs translationally repress hundreds of direct target genes. Using an unbiased approach, we identified the seed sequence in the 3' UTR as a primary motif of miRNA-mediated regulation of protein production. The seed correlated with both mRNA degradation and translational repression (Fig. 4c, d).

Perhaps surprisingly, the repressive effect on individual proteins was relatively small and rarely exceeded fourfold. Because we performed pulsed labelling, this result cannot be explained by persistence of stable proteins. Other investigators observed much higher fold changes (up to 30-fold) in a similar system (double-stranded RNA (dsRNA) transfection in HeLa cells) with artificial reporter constructs⁴¹. One explanation for this apparent discrepancy is that very few ($<0.5\%$) 3' UTRs in our data set have more than three seed sites for a given miRNA (and this value is representative for the whole genome) whereas artificial reporter constructs are designed to contain up to six closely spaced miRNA binding sites.

Identifying functionally important miRNA targets is crucial for understanding miRNA functions. By directly measuring changes in protein production, pSILAC data are likely to be more relevant to the phenotypes than microarray data. We also note that a number of targets are almost exclusively repressed at the level of translation and hence missed by microarrays. pSILAC allows assessment of the early effects of miRNAs on translation. This is a considerable advantage over techniques that assay changes in steady-state protein levels and are therefore almost certainly confounded by indirect effects. Although not all changes in peptide peak intensities reflect true differences in protein synthesis, a direct comparison of pSILAC and luciferase measurements yields very similar results over two orders of magnitude²⁶. Catalogues of proteotypic peptides will further improve this accuracy and help to achieve full-proteome coverage⁴⁵. pSILAC and microarray data can be queried at <http://psilac.mdc-berlin.de>.

Although artificially overexpressing miRNAs might cause non-physiological effects, we found that overexpression and knockdown

of let-7b inversely modulates protein production, suggesting that such effects do not dominate. Nevertheless, transfecting miRNAs that are not endogenously expressed will probably expose many mRNAs to miRNAs that are never coexpressed in the same cell type. Therefore, it could be argued that a number of targets for miR-1 and miR-155 identified by pSILAC are irrelevant *in vivo*. However, transfecting a tissue-specific miRNA into HeLa cells shifts the entire gene expression profile towards that tissue¹². Furthermore, we show that evolutionarily conserved target sites cause stronger effects than non-conserved sites. Altogether, our data probably contain many physiologically relevant direct targets. These arguments are strengthened by the highly significant correlation of pSILAC data with a number of published miRNA target predictions. Seed-based methods had the highest overlap with pSILAC data. Consistently, many down-regulated genes could be explained by seed sites. A number of repressed proteins without seeds are nevertheless probably direct targets of the respective miRNAs. However, although some algorithms include searches for such sites, it seems that they could not identify these non-canonical sites with high success.

Our data indicate that most targets are repressed at both the mRNA and the translational level. As revealed by Fig. 4d, how much both processes contribute to downregulation depends on the individual miRNA-mRNA pair. To test whether targets with strong translational repression share functional properties, we performed gene ontology analysis for proteins with large protein and mRNA fold-change differences (\log_2 -fold change pSILAC – mRNA < -0.3). Intriguingly, we found over-representation of proteins synthesized at endoplasmic-reticulum-associated ribosomes (gene ontology categories 'intrinsic to membrane' and 'endoplasmic reticulum', corrected *P*-values < 0.0001 and < 0.005 , respectively; Supplementary Table 2). Hence, translational repression seems stronger for mRNAs translated at endoplasmic-reticulum-associated ribosomes compared to free cytosolic ribosomes. Thus, endoplasmic-reticulum-associated ribosomes might be more sensitive to miRNA-mediated translational repression. It is tempting to speculate that mRNAs from free ribosomes but not from endoplasmic-reticulum-associated ribosomes are targeted to processing bodies (P-bodies) for degradation⁴⁶. Because the endoplasmic reticulum is considered to lack proteolytic activity, this finding also suggests that co-translational degradation of nascent peptides is not the predominant mechanism of miRNA-mediated translational repression for this subset of targets⁴⁷.

Finally, we showed that overexpression and knockdown of let-7b had largely inverse effects on the protein production of thousands of genes, indicating that altering stationary levels of an endogenously expressed miRNA can tune synthesis levels of a major fraction of the proteome. We noticed that Dicer, which has several let-7 3' UTR seeds, is one of the most strongly upregulated genes in the let-7b knockdown pSILAC (>4 -fold) but not in the microarray data

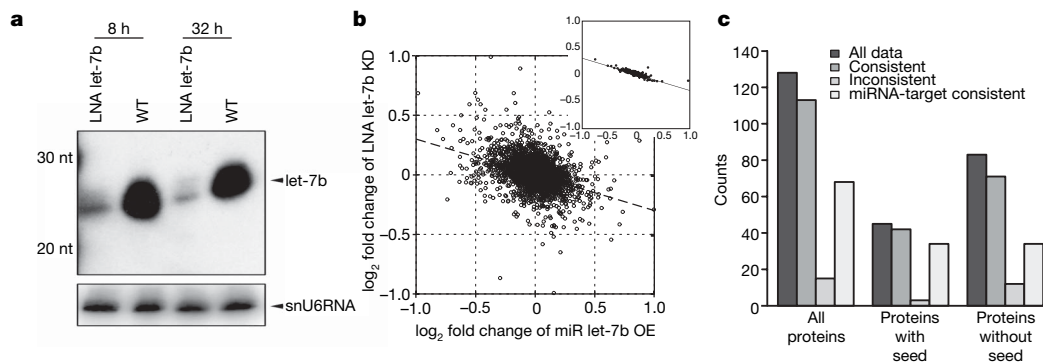


Figure 5 | Endogenous miRNA knockdown. **a**, Northern blotting demonstrates specific and stable let-7b knockdown by means of LNA. nt, nucleotide; WT, wild type. **b**, Scatter plot of changes in protein production in the let-7b overexpression (OE) versus the let-7b knockdown (KD) experiments. The inset shows the same data, averaged over bins of 20 genes.

c, 'Consistent' refers to proteins with pSILAC fold changes that were upregulated in one experiment but down in the other, and 'inconsistent' refers to all other cases. 'miRNA-target consistent' is the subset of 'consistent' proteins that were downregulated in the overexpression experiment but upregulated in the knockdown.

(<1.3-fold). Therefore, Dicer is likely to be a direct translational target of let-7b. This raises the interesting possibility that let-7b regulates mature miRNA levels, which may in part explain our findings.

METHODS SUMMARY

HeLa cells were transfected with 100 nM synthetic dsRNAs designed to mimic mature endogenous miRNAs using DharmaFECT1 (Dharmacon) at 60–70% confluence, or with LNA-anti-let-7b (BioTez). Mock transfections were performed in the same way but without miRNAs. Eight hours post-transfection, cells were split into new dishes containing medium-heavy and heavy SILAC medium prepared as described⁴⁸ and incubated for 24 h until harvest. Corresponding protein and mRNA samples were always derived from the same transfection experiment. For the proteome analysis, miRNA/LNA-transfected cells and corresponding control cells were combined, lysed, and separated by SDS–PAGE. Gel lanes were cut into 15 slices, reduced, alkylated and trypsin-digested. Peptides were extracted and analysed by liquid chromatography–tandem mass spectrometry on a LTQ–Orbitrap hybrid mass spectrometer (Thermo Fisher). All samples were analysed in triplicate resulting in 45 mass spectrometry runs (5 days measurement time) per sample. Raw data files were processed with MaxQuant developed by J. Cox and M. Mann at the Max Planck Institute of Biochemistry (personal communication). False discovery rates were estimated using the target-decoy strategy⁴⁹ against an in-house-curated version of the IPI human protein database (version 3.37). In total, we identified 3,097,418 peptides (66,989 unique sequences) with average absolute mass accuracy of 0.65 p.p.m. We identified 4,961 unique proteins with at least two peptides each at a maximum false discovery rate of 1%. In individual experiments we only considered protein quantifications based on at least three peptide quantifications. Microarray analyses were performed with Human Genome U133 Plus 2.0 chips (Affymetrix), normalized by the standard rma() function (<http://www.bioconductor.org>) and annotated with the current NetAffx annotation file (<http://www.affymetrix.com>).

Received 8 April; accepted 3 July 2008.

Published online 30 July 2008.

- Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Bushati, N. & Cohen, S. M. microRNA functions. *Annu. Rev. Cell Dev. Biol.* **23**, 175–205 (2007).
- He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.* **5**, 522–531 (2004).
- Lai, E. C. miRNAs: whys and wherefores of miRNA-mediated regulation. *Curr. Biol.* **15**, R458–R460 (2005).
- Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Rev. Genet.* **9**, 102–114 (2008).
- Shyu, A. B., Wilkinson, M. F. & van Hoof, A. Messenger RNA regulation: to translate or to degrade. *EMBO J.* **27**, 471–481 (2008).
- Mathonnet, G. et al. MicroRNA inhibition of translation initiation *in vitro* by targeting the cap-binding complex eIF4F. *Science* **317**, 1764–1767 (2007).
- Thermann, R. & Hentze, M. W. *Drosophila* miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature* **447**, 875–878 (2007).
- Wakiyama, M., Takimoto, K., Ohara, O. & Yokoyama, S. Let-7 microRNA-mediated mRNA deadenylation and translational repression in a mammalian cell-free system. *Genes Dev.* **21**, 1857–1862 (2007).
- Liu, J. Control of protein synthesis and mRNA degradation by microRNAs. *Curr. Opin. Cell Biol.* **20**, 214–221 (2008).
- Lim, L. P. et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
- Bentwich, I. Prediction and validation of microRNAs and their targets. *FEBS Lett.* **579**, 5904–5910 (2005).
- Hofacker, I. L. How microRNAs choose their targets. *Nature Genet.* **39**, 1191–1192 (2007).
- Rajewsky, N. microRNA target predictions in animals. *Nature Genet.* **38** (Suppl), S8–S13 (2006).
- Sethupathy, P., Megraw, M. & Hatzigeorgiou, A. G. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nature Methods* **3**, 881–886 (2006).
- Beitzinger, M. et al. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.* **4**, 76–84 (2007).
- Easow, G., Teleman, A. A. & Cohen, S. M. Isolation of microRNA targets by miRNP immunoprecipitation. *RNA* **13**, 1198–1204 (2007).
- Karginov, F. V. et al. A biochemical approach to identifying microRNA targets. *Proc. Natl Acad. Sci. USA* **104**, 19291–19296 (2007).
- Hendrickson, D. G. et al. Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. *PLoS ONE* **3**, e2126 (2008).
- Aleman, L. M., Doench, J. & Sharp, P. A. Comparison of siRNA-induced off-target RNA and protein effects. *RNA* **13**, 385–395 (2007).
- Vinther, J. et al. Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucleic Acids Res.* **34**, e107 (2006).
- Stefani, G. & Slack, F. J. Small non-coding RNAs in animal development. *Nature Rev. Mol. Cell Biol.* **9**, 219–230 (2008).
- Mann, M. Functional and quantitative proteomics using SILAC. *Nature Rev. Mol. Cell Biol.* **7**, 952–958 (2006).
- Ong, S. E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
- Schwanhaeusser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* (in the press).
- Milner, E., Barnea, E., Beer, I. & Admon, A. The turnover kinetics of major histocompatibility complex peptides of human cancer cells. *Mol. Cell. Proteomics* **5**, 357–365 (2006).
- Pratt, J. M. et al. Dynamics of protein turnover, a missing dimension in proteomics. *Mol. Cell. Proteomics* **1**, 579–591 (2002).
- Lam, Y. W., Lamond, A. I., Mann, M. & Andersen, J. S. Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Curr. Biol.* **17**, 749–760 (2007).
- Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130**, 395–398 (2007).
- Cravatt, B. F., Simon, G. M. & Yates, J. R. III. The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000 (2007).
- Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006).
- Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401–1414 (2007).
- Sood, P. et al. Cell-type-specific signatures of microRNAs on target mRNA expression. *Proc. Natl Acad. Sci. USA* **103**, 2746–2751 (2006).
- Nielsen, C. B. et al. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**, 1894–1910 (2007).
- Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
- Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature Struct. Mol. Biol.* **11**, 599–606 (2004).
- Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
- Schwarz, D. S. et al. Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLoS Genet.* **2**, e140 (2006).
- Vasudevan, S., Tong, Y. & Steitz, J. A. Switching from repression to activation: microRNAs can up-regulate translation. *Science* **318**, 1931–1934 (2007).
- Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev.* **17**, 438–442 (2003).
- Stenvang, J. et al. The utility of LNA in microRNA-based cancer diagnostics and therapeutics. *Semin. Cancer Biol.* **18**, 89–102 (2008).
- Wahlestedt, C. et al. Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proc. Natl Acad. Sci. USA* **97**, 5633–5638 (2000).
- Boyerinas, B. et al. Identification of let-7-regulated oncofetal genes. *Cancer Res.* **68**, 2587–2591 (2008).
- Kuster, B., Schirle, M., Mallick, P. & Aebersold, R. Scoring proteomes with proteotypic peptide probes. *Nature Rev. Mol. Cell Biol.* **6**, 577–583 (2005).
- Eulalio, A., Behm-Ansmant, I. & Izaurralde, E. P bodies: at the crossroads of post-transcriptional pathways. *Nature Rev. Mol. Cell Biol.* **8**, 9–22 (2007).
- Nottrott, S., Simard, M. J. & Richter, J. D. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nature Struct. Mol. Biol.* **13**, 1108–1114 (2006).
- Ong, S. E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols* **1**, 2650–2660 (2007).
- Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4**, 207–214 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Cox and M. Mann for early access to the MaxQuant software package, N. D. Socci for discussions, S. Schmidt, G. Born and N. Huebner for the hybridizations at the MDC microarray facility, C. Sommer for technical assistance, M. Huska and M. Andrade-Navarro for setting up the pSILAC website, P. Sharp for a CXCR4 luciferase construct, M. Peter for the IMP-1 reporters, and the Bundesministerium für Bildung und Forschung for funding mass spectrometry instrumentation. R.K. gratefully acknowledges a DAAD scholarship for research stays at the MDC. pSILAC and microarray data can be queried at <http://psilac.mdc-berlin.de>.

Author Contributions M.S. and N.R. conceived, designed and supervised the experiments. B.S. and N.T. performed the wet lab experiments. M.S., Z.F., R.K. and N.R. analysed genome-wide data. M.S., R.K. and N.R. interpreted the data. M.S. and N.R. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.S. (matthias.selbach@mdc-berlin.de) and N.R. (rajewsky@mdc-berlin.de).

ARTICLES

The impact of microRNAs on protein output

Daehyun Baek^{1,2*}, Judit Villén^{3*}, Chanseok Shin^{1,2*}, Fernando D. Camargo¹, Steven P. Gygi³ & David P. Bartel^{1,2}

MicroRNAs are endogenous ~23-nucleotide RNAs that can pair to sites in the messenger RNAs of protein-coding genes to downregulate the expression from these messages. MicroRNAs are known to influence the evolution and stability of many mRNAs, but their global impact on protein output had not been examined. Here we use quantitative mass spectrometry to measure the response of thousands of proteins after introducing microRNAs into cultured cells and after deleting *mir-223* in mouse neutrophils. The identities of the responsive proteins indicate that targeting is primarily through seed-matched sites located within favourable predicted contexts in 3' untranslated regions. Hundreds of genes were directly repressed, albeit each to a modest degree, by individual microRNAs. Although some targets were repressed without detectable changes in mRNA levels, those translationally repressed by more than a third also displayed detectable mRNA destabilization, and, for the more highly repressed targets, mRNA destabilization usually comprised the major component of repression. The impact of microRNAs on the proteome indicated that for most interactions microRNAs act as rheostats to make fine-scale adjustments to protein output.

Large-scale approaches for studying the regulatory effects of microRNAs (miRNAs) have revealed important insights into target recognition and function. These approaches include computational analysis of the selective maintenance or avoidance of miRNA complementary sites during evolution^{1–8} and experimental identification of messages destabilized or those preferentially associated with argonaute proteins in the presence of a miRNA^{7–15}. Despite their utility, none of these approaches directly measures the influence of a miRNA on protein output, which is the most relevant readout of its regulatory effects. The influence of miRNAs on protein output has instead been limited to single-protein analyses, primarily immunoblotting and reporter assays, and a medium-size proteomics analysis with detection of 504 proteins¹⁶.

Proteomic consequences of added miRNAs

To acquire data sufficient to investigate the effects of miRNA regulation on the proteome, we applied a quantitative-mass-spectrometry-based approach using SILAC (stable isotope labelling with amino acids in cell culture)¹⁷ to investigate the influence of specific miRNAs on the levels of many proteins (Supplementary Figs 1 and 2). We first measured the effects of introducing miR-124, a brain-specific miRNA, into HeLa cells. To include proteins from a broad expression spectrum, this experiment focused on nuclear-localized proteins. Out of 2,120 proteins detected, the analysis considered 1,544 that mapped to our non-redundant mRNA data set and were each quantified by at least two independent measurements that passed our quality thresholds (Supplementary Data 1 and 5).

Because this and all subsequent SILAC analyses were performed with two technical replicates, and because different peptides from the same protein and different charge states from the same peptide also provided the opportunity for independent measurements, most proteins were quantified by many more than two independent measurements (median of 12 for the 1,544 quantified proteins). The high

reproducibility when comparing technical replicates and when comparing different peptides representing the same protein illustrated the quantification accuracy ($r^2 = 0.72$ and 0.65 , respectively, Spearman's correlation; Supplementary Fig. 3).

Messages for proteins that decreased the most relative to the mock-transfection control were compared to the messages of the other quantified proteins (cutoff, 85th percentile), searching for motifs over-represented in their open reading frames (ORFs) or untranslated regions (UTRs). When considering all 16,384 possible 7-nucleotide motifs and the different regions of the mRNA, the only one significantly enriched after Bonferroni correction for multiple hypothesis testing was the GUGCCUU heptanucleotide in the 3' UTR ($P < 10^{-7}$, Fisher's exact test). This heptanucleotide motif comprised the 6-nucleotide match to the seed of miR-124 (underlined) supplemented by a match to miRNA nucleotide 8, and is named the 7mer-m8 seed-matched site (Fig. 1a). It was the same motif that is most associated with 3' UTRs of messages destabilized after introduction of miR-124 (ref. 9). The other sites consistently associated both with preferential conservation and with mRNA destabilization after miRNA introduction are named the 6mer, 7mer-A1 and 8mer seed-matched sites^{2,7,8} (Fig. 1a). A more directed search for the seed-matched sites revealed that most of the robustly repressed proteins derived from messages with at least one 7–8mer 3'-UTR site (Fig. 1b). For example, 24 out of the 40 proteins repressed by at least 50% had at least one 7–8mer 3'-UTR site, with only 3 of these 24 attributed to chance (Fig. 1b, repression cutoff of 50%). Less stringent repression cutoffs yielded many additional proteins from messages with 7–8mer sites, even after subtracting those expected by chance. The overall enrichment of seed-matched sites in messages of downregulated proteins indicated that miR-124 recognition of mRNAs for repression of protein output used, more than any other type of site, seed-matched sites in 3' UTRs.

To survey the efficacy of the different seed-matched sites, we plotted the response of proteins from messages with 3' UTRs possessing

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Department of Cell Biology, 240 Longwood Avenue, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

single sites (Fig. 1c). Proteins from messages with single 7–8mer sites had a significant propensity to be downregulated when compared to those from messages without 3′-UTR sites ($P = 0.02$, 0.0008 and 0.02 for 8mer, 7mer-m8 and 7mer-A1, respectively, Kolmogorov–Smirnov test).

We performed analogous SILAC experiments with two additional miRNAs: miR-1 and miR-181, for which 2,312 and 1,774 proteins, respectively, mapped to our non-redundant mRNA data set and passed our quantification quality cutoffs (Supplementary Data 2, 3 and 5). The motifs associated with messages of the most downregulated proteins mirrored those observed for miR-124; for miR-1, the 7mer-m8 match was the most confidently enriched heptanucleotide motif in the 3′ UTRs of downregulated proteins ($P = 0.0004$), and, for miR-181, the 7mer-A1 match was among the top two motifs ($P = 0.007$), slightly less confidently enriched than an unrelated motif, CUGCCCC ($P = 0.006$, Fisher's exact test with Bonferroni correction).

When pooling the data from all three miRNA transfections, thereby combining 5,630 independent protein quantifications, proteins from messages with single 7mer or 8mer sites matching the cognate miRNA had a significant propensity to be downregulated (Fig. 1d, $P < 10^{-14}$ overall, $P < 10^{-4}$ for each site separately, Kolmogorov–Smirnov test). Vertical displacement from the no-site distribution demonstrated that at least 16% of the proteins from messages with single 7–8mer 3′-UTR sites responded to the miRNA (Fig. 1d). The response of proteins from messages with a 6mer site closely tracked that from messages with no site, indicating that in this system 6mer recognition was generally insufficient for detectable protein downregulation (Fig. 1d).

Analysis of site conservation, site depletion, argonaute pull-downs and reporter assays all indicate that targeting can occur in protein-coding regions^{2,5,13,18}. Analysis of mRNA destabilization concurs that targeting occurs in coding regions, but indicates that these sites are generally much less effective than those in 3′ UTRs⁷. However, monitoring mRNA destabilization would understate the influence of sites

in coding regions if these sites, by virtue of falling in the path of the ribosome, had a disproportionate effect on translation compared to mRNA destabilization. To address this possibility, we examined our data monitoring protein output and found that sites in coding regions were generally less effective than those in UTRs (Fig. 1e).

Proteomic consequences of disrupting *mir-223*

Measuring the effects of ectopic miRNA addition can provide generic insights into miRNA target recognition, but the responsive proteins are not necessarily the endogenous targets, and the magnitude and kinetics of mRNA and protein changes are not expected to match those of endogenous targeting (Supplementary Discussion). To obtain data relevant to endogenous miRNA–target interactions, with pertinent information on the degree of repression, we examined the effects of the *mir-223* gene knockout in mouse neutrophils. *mir-223* is preferentially expressed in myeloid haematopoietic cells, with high expression in neutrophils and their progenitors^{19,20}. To obtain labelled samples suitable for the quantitative proteomics experiment, we isolated bone marrow haematopoietic progenitors from wild-type and *mir-223*-deficient mice²¹ and developed a protocol for their proliferation in SILAC media and differentiation into mature neutrophils *in vitro* (Fig. 2a and Supplementary Fig. 4a, b). By day 8, the surviving cells had descended from progenitors that had undergone multiple cell divisions in the presence of SILAC media (Supplementary Fig. 4c), which resulted in >99% heavy isotope incorporation. RNA blots confirmed that both the progenitors and the differentiating neutrophils expressed *mir-223* (Fig. 2b). Array experiments demonstrated that the effect of miR-223 on messages with cognate sites was analogous to that observed for neutrophils isolated directly from mice, although somewhat less robust (Fig. 2c), perhaps in part because the neutrophils differentiated *in vitro* accumulate ~35% less miR-223 (Fig. 2b).

Analysis by mass spectrometry of both nuclear and cytoplasmic fractions provided quantitative information for 5,019 proteins, 3,819

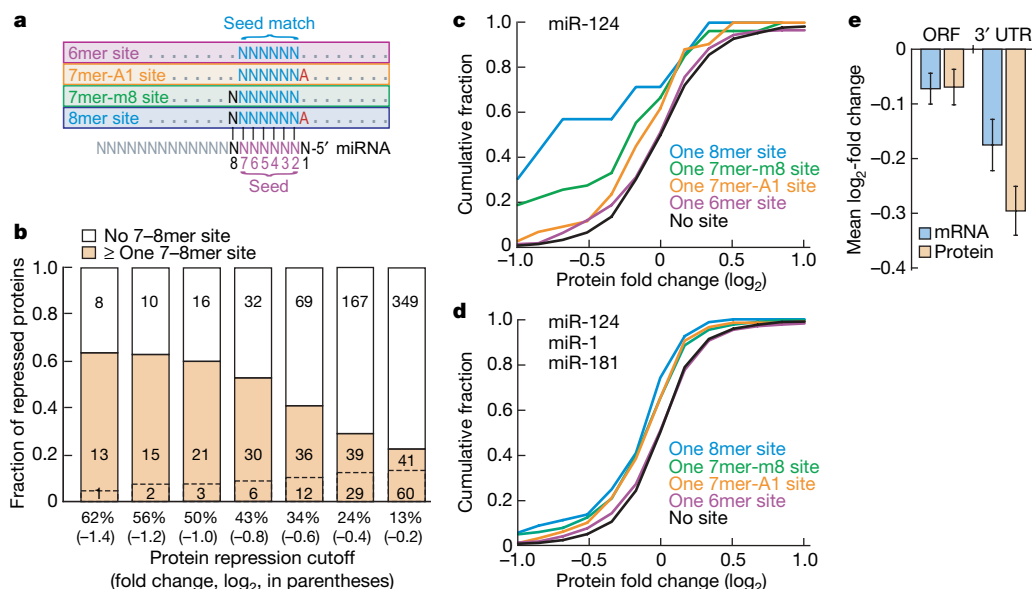


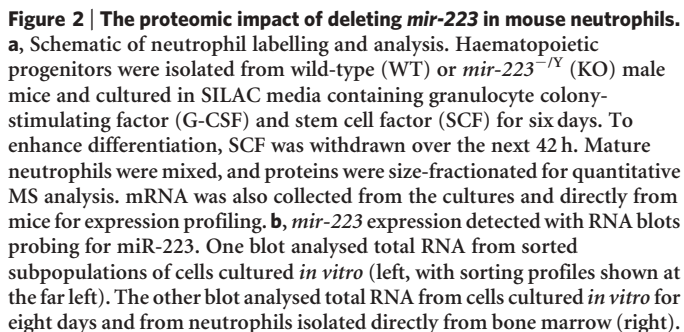
Figure 1 | The impact of transfected miRNAs on protein output.

a, Canonical miRNA seed-matched sites. **b**, The fraction of repressed proteins deriving from messages with miR-124 3′-UTR sites (filled orange bar). At each repression cutoff, the number of repressed proteins from messages without 3′-UTR sites (indicated in the open bar) was used to calculate the additional fraction expected by chance to have a site (dashed line, with the corresponding number of repressed proteins indicated below the dashed line). Above the dashed line is the surplus number of repressed proteins deriving from messages with sites. **c**, Response of proteins from messages with single miR-124 3′-UTR sites. Plotted is the fraction of proteins that change at least to the degree indicated on the x axis. Proteins

from messages with multiple 3′-UTR sites were not considered. 6mer sites that were part of larger sites were not included in the 6mer distribution, and 7mers that were part of 8mers were not included in the 7mer distributions. **d**, Efficacy of single 3′-UTR sites when pooling data from miR-124, miR-1 and miR-181 transfections, plotted as in **c**. **e**, ORF and 3′-UTR targeting efficacy. Plotted is the average change (\pm standard error) of protein and corresponding mRNA for quantified proteins from messages with at least one 8mer in the ORF ($n = 83$) or 3′ UTR ($n = 87$) corresponding to the transfected miRNA (excluding messages with sites in both ORF and 3′ UTR).

expected from each site acting independently. This independent, non-cooperative response was in agreement with results monitoring mRNA destabilization and reporter assays, which indicate that cooperative action of sites tends to occur only for those sites falling within 8–40 nucleotides of each other⁷. Taken together, our results demonstrated experimentally that targeting principles elucidated from ectopically added miRNAs apply also to endogenous miRNA targeting, and in particular to endogenous targeting at the level of protein downregulation.

The perturbation of endogenous targeting provided the opportunity to test sets of target predictions. When considering current predictions from miRBase Targets²², miRanda^{23,24}, PicTar^{4,25}, PITA²⁶ and TargetScan^{2,7}, all of which use site conservation as a prediction criterion, those from TargetScan and PicTar performed the best (Fig. 3a). Predictions from TargetScan and PicTar are primarily those messages with at least one 3'-UTR 7-8mer site conserved among mammals, operationally defined as those sites preserved in orthologous locations of human, mouse, rat and dog UTRs^{2,4}. Their enhanced performance over the set of messages with any 3'-UTR



©2008 Macmillan Publishers Limited. All rights reserved

7–8mer sites demonstrated that considering site conservation not only enriches for sites with presumed functional roles but also enriches for those that are more effective. All of the other algorithms include many sites with least one mismatch or wobble to the seed, which seems to have compromised their performance. For example, the predictions of miRBase Targets had been generated using the miRanda algorithm²³ with updated parameters, searching for conserved sites with more stringent seed pairing but still allowing one mismatch or wobble to the seed²². Analysis of the seed-matched and seed-mismatched predictions separately revealed that any benefit gained in searching for site conservation was offset by the inclusion of many poorly performing predictions with seed mismatches (Supplementary Fig. 5a). Despite the relative success of TargetScan and PicTar, two-thirds of their predicted targets appeared to be non-responsive to miR-223 loss in neutrophils, indicating a false-positive

rate within the range of that inferred from estimates of chance conservation of the target sites of this miRNA (Supplementary Fig. 5b).

The similar performance of PicTar and TargetScan was expected for miR-223, which begins with a U, but might not have been expected for those miRNAs that do not begin with a U. TargetScan rewards an A across from position 1, whereas PicTar (and similar algorithms^{3,27}) rewards a Watson–Crick match at this position. Therefore, for miRNAs that begin with A, C or G, only one of the two heptanucleotide matches (the 7mer-m8) is the same for the algorithms^{2,4} and thus about half of the predicted targets are expected to differ. To investigate which type of heptanucleotide match is most associated with decreased protein output, we examined the proteomics data from the experiment transfecting miR-181, which does not begin with a U. Plotting the response of proteins from messages with single sites revealed that the 7mer-A1 match was more effective than the Watson–Crick 1–7 match (Fig. 3b, $P = 0.009$, Kolmogorov–Smirnov test). Moreover, the Watson–Crick 1–7 match was no more effective than were 6mer sites with G or C mismatches across from position 1 (Fig. 3b). We conclude that the recognition of an A across from miRNA nucleotide 1 favours miRNA-mediated protein down-regulation, which explains the preferential conservation of an A at this position, even when it cannot participate in a Watson–Crick interaction².

Target prediction sets are typically ranked, with the assertion that the better scoring predictions are more likely to be authentic or effective. Recent TargetScan predictions (release 4) are ranked by ‘total context score’, which is based on site type, site number and site context⁷. This ranking correlated with protein downregulation, with the top third significantly more responsive than the bottom third (Fig. 3c). For the other algorithms, the predictions scoring in the top third were not significantly more responsive than those in the bottom third (Fig. 3c, $P > 0.05$, Mann–Whitney U -test). Despite their poor overall performance, the more inclusive algorithms might still have utility when considering only their top few predictions. To investigate this possibility, we considered only the top 29 predictions of each algorithm, choosing 29 because the most restrictive set (that of PicTar) includes this number of predictions. At this stringent cut-off, the performances of the more inclusive algorithms approached that of PicTar (resulting in difference that was no longer statistically significant, $P > 0.05$), but remained lower than that of TargetScan ($P < 0.05$, Fig. 3d). Interestingly, the top 29 quantified proteins ranked only by the total context score of their respective 3' UTRs, without any regard to site conservation, were at least as responsive as the top 29 TargetScan predictions (Fig. 3d).

Analysis of the evolutionary impact of miRNAs and analysis of messages that are upregulated in miRNA-deficient animals both indicate that many non-conserved sites mediate repression *in vivo*^{5,6,10,11}. We also found evidence for widespread non-conserved targeting among natural miR-223–target interactions. In an attempt to predict non-conserved targets, RNA22 (ref. 28) and a more permissive version of PITA²⁶ do not consider site conservation. When evaluated using our miR-223 data, these algorithms performed no better than did a simple search for messages with 7–8mer seed-matched sites (Fig. 3e). A more effective tool was the total context score, which correlated with derepression when considering only those messages with non-conserved 7–8mer sites (that is, sites missing or mutated in orthologous positions of human, rat or dog 3' UTRs), with the top third of non-conserved predictions significantly more effective than the bottom third (Fig. 3e). Indeed the top third of non-conserved predictions (Fig. 3e, context score) appeared as effective as the bottom two-thirds of conserved predictions (Fig. 3c, TargetScan), and because proteins from non-conserved predictions outnumbered those from conserved ones by 6 to 1, the non-conserved predictions with favourable context scores were a bountiful source of biological targets.

The success of the total context score in ranking both conserved and non-conserved predictions was due in part to its consideration of

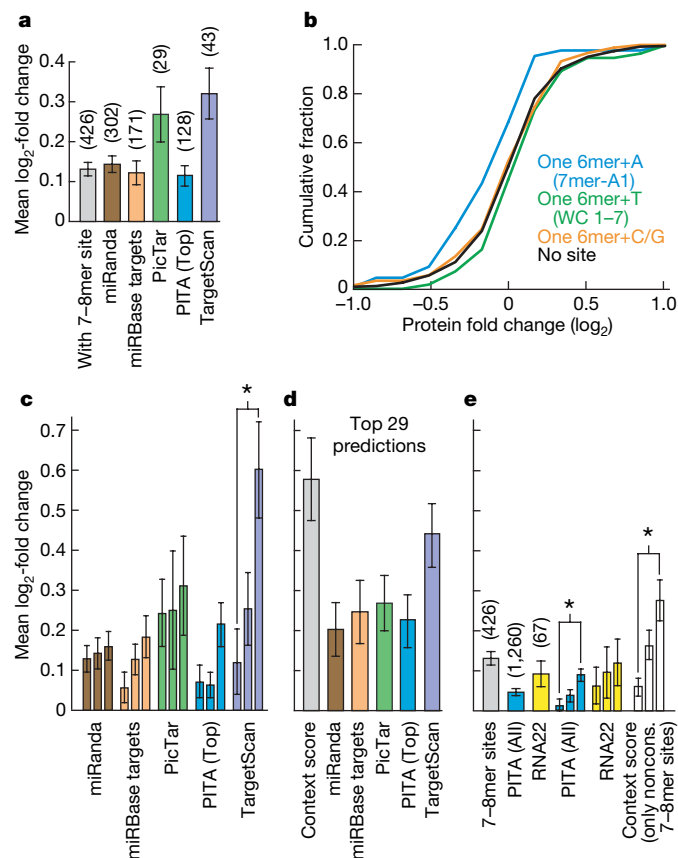


Figure 3 | Correspondence between computational target predictions and observed protein changes. Analogous results were observed using the transfection data sets (Supplementary Fig. 8). **a**, Performance of programs that consider site conservation. Plotted is the average protein derepression (\pm standard error) of genes with ≥ 1 conserved or non-conserved 7–8mer 3'-UTR site (grey) and of genes predicted to be miR-223 targets. The number of quantified proteins in each set is in parenthesis. **b**, Recognition of an adenosine (A) opposite the first nucleotide of the miRNA. The cumulative plot of protein changes after miR-181 transfection compares proteins from messages with no seed-matched 3'-UTR site to those from messages with the indicated single 3'-UTR site. **c**, Relationship between the scores of predicted targets and protein derepression. Predictions corresponding to quantified proteins were divided into three equal-size bins according to the scores proposed to indicate the quality of the prediction or degree of repression. Statistically significant differences between the bottom and top third are indicated (asterisk, $P < 0.01$, Mann–Whitney U -test). **d**, Response of the top 29 predictions of each algorithm, plotted as in **a**. **e**, Performance of programs that do not consider site conservation, displayed as in **a** and **c**. Also shown is the response of quantified proteins from messages with only non-conserved 7–8mer 3'-UTR sites, binned by total context score.

site type (Fig. 2e) and the number of sites (Fig. 2g). To isolate its third component (site context) we considered only those quantified proteins deriving from messages with single 7mer-m8 3'-UTR sites and still observed a significant correlation between context score and protein response ($P = 0.001$, Spearman's correlation test). Predicted 3'-UTR structure and other features of site context are reported to influence site accessibility and efficacy^{7,8,26,27,29–31}. The context score combines some of these features, including high local AU nucleotide composition (which accounts for effects of predicted 3'-UTR structure on site accessibility), proximity to residues that can pair to miRNA nucleotides 13–16, and positioning away from the centre of long UTRs⁷. As anticipated from analyses of mRNA destabilization data⁷, the most influential component was local AU composition, which when examined in isolation significantly correlated with protein response ($P = 0.01$, Spearman's correlation test).

Response of proteins compared to that of mRNAs

Because previously used high-throughput methods were unable to determine the amount of protein repression, the relative contributions of mRNA destabilization and translational repression during miRNA-mediated regulation has been of intense interest. Our miR-223 data was informative for addressing this issue because it examined the response, at both the mRNA and the protein level, of removing an endogenous miRNA, without the confounding influences of exogenous targeting mediated by an ectopically delivered miRNA. The near steady-state nature of our miR-223 system also avoided quantification caveats inherent to transient transfection, such as variable transfection efficiencies and pre-steady-state complexities especially acute when comparing effects on an mRNA to those on its protein because messages and their proteins can have very different intrinsic stabilities. Note that our mRNA quantification used standard array platforms, which include oligo(dT) priming during detection, and thus the

mRNA destabilization we observed encompassed the conversion of the message into a form that was unsuitable for translation because it lacked a poly(A) tail.

To achieve greater quantification accuracy in this analysis of individual proteins, we narrowed our focus to the 2,773 proteins quantified with ≥ 6 independent measurements. Plotting protein changes as a function of mRNA changes indicated a strong positive correlation for messages with 7mer or 8mer 3'-UTR sites (Fig. 4a; $r^2 = 0.45$ and 0.63 , $P < 10^{-33}$ and $P < 10^{-12}$, respectively) and weaker correlation for messages without sites (Fig. 4b; $r^2 = 0.15$, $P < 10^{-11}$, Pearson's correlation test). Proteins in both plots displayed some scatter around the origin; however, when normalizing to those without sites, many more of those from messages with sites increased in response to miR-223 loss (Fig. 4a, b and Supplementary Fig. 6). Immunoblots probing for three of the more responsive proteins confirmed protein derepression in *mir-223*^{-/-} neutrophils differentiated *in vitro* as well as in those isolated directly from mice (Supplementary Fig. 7).

Two of the three most responsive proteins derived from messages with single, non-conserved 7mers (Table 1)—sites that on their own would not be expected to impart such a robust response. Previous work has shown that sites falling within 8–40 nucleotides of sites to co-expressed miRNAs typically act cooperatively, which increases the effect of loosing interactions at particular sites⁷. We performed high-throughput sequencing to identify miRNAs co-expressed in cultured neutrophils (Supplementary Table 2) and found that both of the highly responsive 7mers fell near to sites matching a co-expressed miRNA, with intersite spacing favouring a cooperative response (Table 1). The site in *Ctsl* was near a site for the miR-26 family, one of five families sequenced more frequently than miR-223, whereas the site in *Gns* fell near a site to the miR-103/107 family, sequenced about a third as often as miR-223 (Supplementary Table 2).

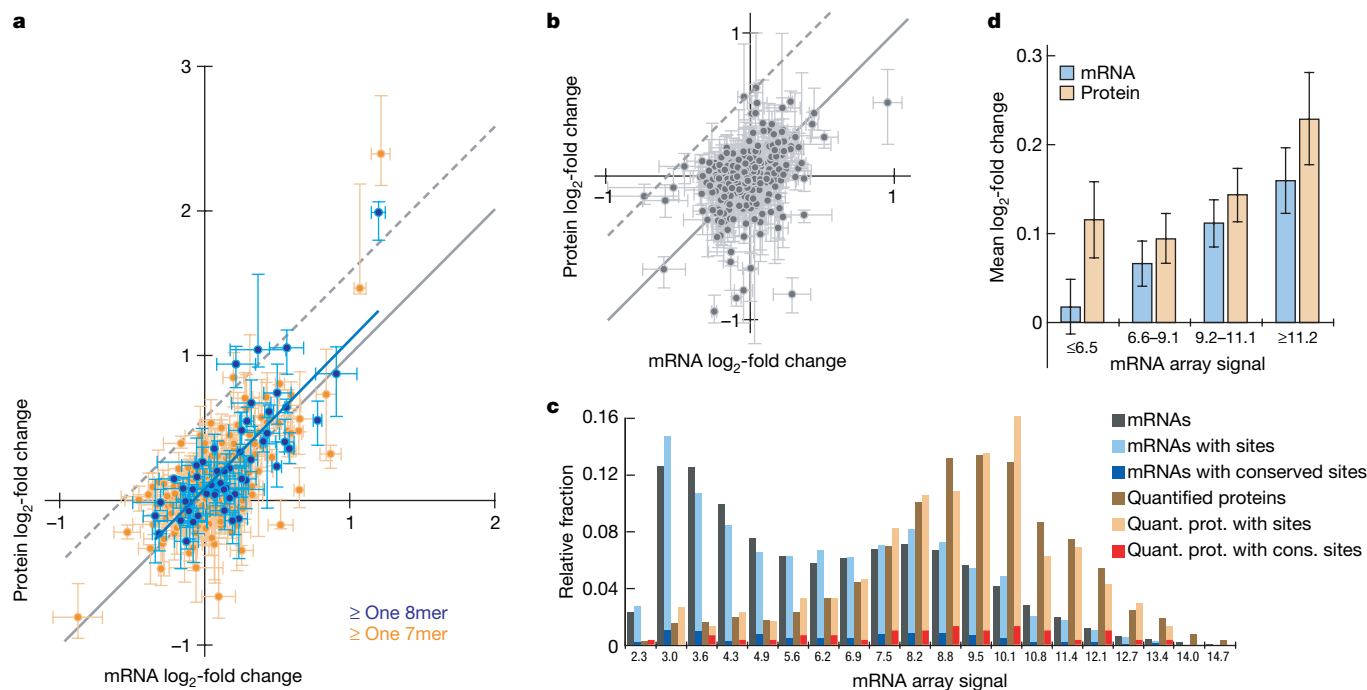


Figure 4 | Comparison of protein and mRNA changes accompanying miR-223 loss. **a**, Protein and mRNA changes for quantified proteins deriving from messages with at least one 8mer 3'-UTR site (blue, $n = 55$) or at least one 7mer (orange, an additional 250 proteins). The least-squares best fit to the 8mer data are shown (blue line), as are reference lines (grey), which both have slope of 1.0. Vertical error bars indicate 25th and 75th percentiles for independent measurements of protein changes. Horizontal error bars indicate standard errors of mRNA changes from three biological replicates, one of which was also used for the SILAC experiments. **b**, Protein and mRNA

changes for quantified proteins deriving from messages without 7–8mer 3'-UTR sites, plotted as in **a**. One of seven random cohorts is plotted here; the other six are in Supplementary Fig. 6. **c**, Distribution of the indicated reference-set mRNAs and quantified proteins with respect to mRNA expression, as indicated by the array signals from cultured neutrophils. **d**, Response of quantified proteins and their respective mRNAs to *mir-223* deletion, considering those messages with 7–8mer 3'-UTR sites, grouped by mRNA expression.

If protein changes merely reflected mRNA changes, with no additional repression at the translational level, then the points would fall on the diagonal (Fig. 4a, grey line). Although many were on the diagonal or very close to it, least-squares linear regression yielded a positive y -intercept (+0.053 and +0.079 for 7mer and 8mer data, respectively). These modest yet statistically significant positive y -intercept values ($P = 0.0002$ and $P = 0.042$, t -test) suggested that a cohort of genes were modestly derepressed at the protein level with little or no change at the mRNA level. The messages of such genes were each good candidates for targets affected only at the translational level, although some might have derived from genes undergoing non-miRNA-mediated transcriptional repression as a compensatory feedback response to the loss of miR-223 targeting.

Despite evidence for some translation-only repression, all proteins derepressed by more than 50% ($\log_2 > 0.58$) derived from messages that displayed detectable increases (Fig. 4a and Table 1). Moreover, only five points were more than 0.58 units (\log_2) above the diagonal (Fig. 4a, upper dashed line; Table 1, indicated with §). Note that a 33% repression by miR-223 in wild-type neutrophils would correspond to a 50% (+0.58 \log_2) derepression in mutant neutrophils. Thus, in wild-type neutrophils only 5 of the 305 quantified proteins from messages with 7–8mer 3'-UTR sites appeared to undergo translational repression by more than 33%. We conclude that, although in some instances translational repression produces a substantial amount of endogenous miRNA-mediated repression, this occurred for surprisingly few of the many inferred targets. Substantial translational repression appeared so rarely because targets repressed only at the level of translation were repressed quite modestly (<33%); for targets undergoing more robust repression, the major component of the repression was usually mRNA destabilization (Table 1). Further study is required to determine whether those mRNA molecules undergoing miRNA-mediated repression might experience translational repression as a prelude to destabilization, but our results show that mRNA destabilization can explain most of the endogenous miR-223-mediated repression.

Our proteomics data were limited to the confidently quantified proteins, which were expected to be those that were both soluble and more highly expressed in neutrophils. To consider how the expression bias might have influenced our results, we plotted the

distributions of mRNAs and quantified proteins as a function of mRNA expression in neutrophils, considering all mRNAs of our non-redundant data set (including those without detectable expression), as well as those with 3'-UTR sites (Fig. 4c). The messages with conserved or non-conserved 3'-UTR sites displayed the full range of expression values, with a distribution matching that of messages more generally. As anticipated, more quantified proteins derived from highly expressed messages (Fig. 4c). However, the distribution of quantified proteins from messages with sites (conserved or non-conserved) closely matched that of those without sites. Moreover, we found no evidence that the greater representation of proteins from more highly expressed messages underrepresented the impact of miRNAs on protein output; if anything, proteins from more highly expressed messages tended to respond more robustly than did those from lowly expressed messages (Fig. 4d). An analysis using Gene Ontology terms³² came to similar conclusions (data not shown). Therefore, although our experiment monitored the impact on only a portion of the neutrophil proteome and thus missed many miR-223 targets (including some conserved targets, such as Mef2c; refs 2, 21), we found no reason to suspect that undetected targets respond more robustly.

The proteins from the least abundant mRNAs appeared to respond without detectable mRNA changes (Fig. 4d, ≤ 6.5 bin). Apparent dominance of the translational component might have been a consequence of less reliable array signals for these messages, many of which fell within background signals from non-expressed messages. A more intriguing possibility is that very efficient translation of these messages (inferred from the ability to quantify proteins from such lowly expressed messages) makes them more susceptible to greater translational repression.

The regulatory function of miR-223

Some of the most strongly derepressed proteins from messages with miR-223 sites provided potential explanations for the pro-inflammatory phenotype observed in *mir-223*^{-/-} neutrophils²¹. Cathepsin L and cathepsin Z (Ctsl and Ctsz, listed first and fourteenth in Table 1) are cysteine proteases associated with chronic inflammatory conditions, in which they can act as mediators of tissue destruction^{33,34}. Another potentially relevant target, the insulin-like growth factor

Table 1 | The most responsive proteins deriving from messages with at least one 7–8mer 3'-UTR site

Protein*	Fold change (log ₂) in <i>mir-223</i> ^{-/-} cells versus wild type				Fold change (log ₂) of mRNA during neutrophil differentiation†	3'-UTR sites	Co-expressed miRNA family‡ with cooperatively spaced site
	Neutrophil culture (8 days)		Sorted cultured cells				
	Protein (25th–75th percentiles)	mRNA (±s.e.m.)	Progenitor mRNA	Neutrophil mRNA			
Ctsl§	2.40 (2.18–2.80)	1.21 ± 0.07	0.79	1.03	1.71	7mer	miR-26, 8mer
Parp9§	1.99 (1.80–2.06)	1.20 ± 0.05	0.43	1.07	1.55	8mer	
Gns	1.47 (1.43–2.19)	1.07 ± 0.04	0.51	1.01	1.30	7mer	miR-103/107, 7mer
Rasa1	1.06 (0.87–1.18)	0.56 ± 0.12	0.27	0.62	-0.22	8mer	
Acs1§	1.04 (0.92–1.56)	0.37 ± 0.19	0.44	0.50	-1.49	8mer ¶	
Igf1r§	0.94 (0.78–1.06)	0.22 ± 0.13	0.01	-0.03	0.03	8mer , 7mer	
Galnt7	0.87 (0.58–1.06)	0.91 ± 0.14	0.40	0.89	-0.15	8mer, 7mer	miR-103/107, 7mer
Myo1c§	0.85 (0.64–1.02)	0.20 ± 0.07	0.40	0.29	0.55	7mer	
Gm885	0.81 (0.76–0.88)	0.52 ± 0.03	0.07	0.27	1.63	7mer	
Smarcd1	0.74 (0.53–0.94)	0.50 ± 0.09	0.24	0.41	-0.11	8mer	
Mvp	0.73 (0.63–0.82)	0.60 ± 0.01	0.25	0.48	2.07	7mer	
1110019N10Rik	0.73 (0.28–1.05)	0.84 ± 0.07	0.66	0.88	-0.61	7mer	
Ipo9	0.71 (0.65–0.88)	0.28 ± 0.10	0.19	0.33	-0.72	7mer, 7mer	
Ctsz	0.71 (0.65–0.81)	0.42 ± 0.09	0.41	0.70	-0.26	7mer	miR-27, 7mer
Atp2b1	0.67 (0.51–0.83)	0.32 ± 0.14	-0.01	-0.01	0.71	8mer	
Prkcb1	0.66 (0.54–0.76)	0.57 ± 0.09	-0.47	0.37	2.97	7mer, 7mer	
Rrm2	0.64 (0.61–0.70)	0.56 ± 0.01	0.30	0.60	-0.49	8mer	miR-27, 7mer
Ankrd13a	0.64 (0.58–0.87)	0.29 ± 0.08	0.10	0.23	0.05	7mer	
Ywhah	0.61 (0.39–0.66)	0.44 ± 0.03	0.14	0.56	0.33	8mer	miR-142-5p, 7mer

*Listed are all proteins quantified using ≥ 6 independent measurements and also upregulated more than 1.5-fold (0.58 \log_2) in *mir-223*^{-/-} neutrophils.

†mRNA change when comparing sorted *mir-223*^{-/-} neutrophils (cultured for eight days) with sorted *mir-223*^{-/-} progenitors (cultured for four days; Fig. 2b). For messages of all quantified proteins, median fold change (\log_2) was -0.22.

‡Considered were sites for 20 co-expressed miRNA families (Supplementary Table 2).

§Protein upregulated $\geq 50\%$ (0.58 \log_2) with miR-223 loss, after accounting for the mRNA change.

||Conserved site.

¶Conserved as the 8mer in human and rat but as a 7mer in dog.

receptor 1 (Igflr, listed sixth), is crucial for the priming and activation of mature neutrophils^{35,36}.

To examine whether repression begins before neutrophil maturation, we profiled mRNA levels in sorted progenitors and neutrophils (Supplementary Data 4). Messages of most of the highly responsive proteins were derepressed already at the progenitor stage, although usually to a lower degree than in neutrophils, which accumulate more miR-223 (Table 1 and Fig. 2b).

The profiles of miR-223-deficient progenitors and neutrophils provided the opportunity to examine the regulation of putative miR-223 targets in the absence of miR-223 to determine whether miR-223-mediated repression predominantly acts coherently with (that is, in the same direction as) the other gene-regulatory processes acting on these genes. During differentiation from progenitor to neutrophil, putative targets increased and decreased in similar numbers (Table 1 and Supplementary Data 4). This result revealed a proportion of incoherent regulatory relationships larger than that observed for other miRNAs^{5,6,11} but nonetheless consistent with the miR-223 loss-of-function phenotype; this phenotype indicates that miR-223 dampens progenitor proliferation and neutrophil differentiation and activation²¹—functions opposite of those expected for coherent regulatory interactions involving a miRNA preferentially expressed in neutrophils.

Because the miR-223 proteomics experiment detected targeting potentially missed by other high-throughput methods, particularly non-conserved targets influenced (albeit modestly) at the level of translation, it provided the clearest picture so far of the scope and magnitude of endogenous miRNA targeting. The vertical displacement from the no-site distribution in Fig. 2e indicated that at least 18.4% of the 426 proteins from messages with 7–8mer 3′-UTR seed-matched sites underwent increased protein output attributable to the sites, thereby implicating messages for at least 78 out of the 3,819 quantified proteins as direct targets. These 78 included ~33% of those quantified proteins from messages with conserved 3′-UTR sites and ~16% of those from messages with nonconserved 3′-UTR sites. Assuming that only about one-third of the proteome was quantified, we estimate that miR-223 has >200 targets in neutrophils (3 × 78). These would not include any targets undergoing fail-safe regulation (targeting of messages for proteins not normally expressed at all in neutrophils), which are invisible in derepression experiments. Despite the broad scope of miR-223 targeting, each interaction had only a modest effect, even when observed at the protein level. Many miR-223-responsive targets also have sites for other miRNAs, some of which are also expressed in neutrophils, and thus the aggregate impact of miRNAs on these targets is presumably greater than that observed for miR-223 alone. Nonetheless, the targeting by other miRNAs is not expected to obscure the effect of removing miR-223 because multiple non-overlapping sites to co-expressed miRNAs typically act independently^{7,8}, and in the rare cases in which they do not act independently, they act cooperatively, which would boost rather than decrease the effect of losing a single miRNA⁷. The widespread scope but low magnitude of endogenous miR-223-mediated repression indicates that this miRNA often acts as a rheostat to adjust protein output.

METHODS SUMMARY

HeLa cells were grown in media containing either regular (light) Lys and Arg or ¹³C₆-labelled (heavy) Lys and Arg. Light cells were transfected with miRNA, and heavy cells were mock-transfected. After 24 h some cells were harvested for mRNA expression profiling. After 48 h the remaining cells were harvested, and equal numbers from both populations were mixed and enriched for soluble nuclear proteins. Neutrophil culture was as outlined in Fig. 2a. Protein mixtures were separated by SDS-PAGE, and fractions were digested with trypsin. Peptides were analysed by liquid chromatography–tandem mass spectrometry (LC–MS/MS), which identified peptides and quantified the relative amounts of isotopic pairs of the same peptide. To prevent double-counting of any targeting interactions, peptides were mapped to a non-redundant complementary DNA data set (Supplementary Data 5), and targeting analyses were as performed previously on

mRNA destabilization data⁷. To compare to target-prediction algorithms, predictions by TargetScan (release 4.1)^{2,7}, PicTar (human, chimp, mouse, rat, dog)^{4,25}, miRanda (January 2008 release)^{23,24}, miRBase Targets (version 5)²², RNA22 (ref. 28) and PITA²⁶ were obtained from their respective websites, using the most recent predictions publicly available as of March 2008.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 April; accepted 10 July 2008.

Published online 30 July 2008.

- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
- Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
- Brennecke, J., Stark, A., Russell, R. B. & Cohen, S. M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
- Krek, A. et al. Combinatorial microRNA target predictions. *Nature Genet.* **37**, 495–500 (2005).
- Farh, K. K. et al. The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* **310**, 1817–1821 (2005).
- Stark, A., Brennecke, J., Bushati, N., Russell, R. B. & Cohen, S. M. Animal microRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* **123**, 1133–1146 (2005).
- Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
- Nielsen, C. B. et al. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* **13**, 1894–1910 (2007).
- Lim, L. P. et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
- Krutzfeldt, J. et al. Silencing of microRNAs *in vivo* with ‘antagomirs’. *Nature* **438**, 685–689 (2005).
- Giraldez, A. J. et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
- Beitzinger, M., Peters, L., Zhu, J. Y., Kremmer, E. & Meister, G. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.* **4**, 76–84 (2007).
- Easow, G., Teleman, A. A. & Cohen, S. M. Isolation of microRNA targets by miRNP immunoprecipitation. *RNA* **13**, 1198–1204 (2007).
- Karginov, F. V. et al. A biochemical approach to identifying microRNA targets. *Proc. Natl Acad. Sci. USA* **104**, 19291–19296 (2007).
- Zhang, L. et al. Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol. Cell* **28**, 598–613 (2007).
- Vinther, J., Hedegaard, M. M., Gardner, P. P., Andersen, J. S. & Arctander, P. Identification of miRNA targets with stable isotope labeling by amino acids in cell culture. *Nucleic Acids Res.* **34**, e107 (2006).
- Ong, S. E. et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**, 376–386 (2002).
- Kloosterman, W. P., Wienholds, E., Ketting, R. F. & Plasterk, R. H. Substrate requirements for *let-7* function in the developing zebrafish embryo. *Nucleic Acids Res.* **32**, 6284–6291 (2004).
- Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs modulate hematopoietic lineage differentiation. *Science* **303**, 83–86 (2004).
- Fazi, F. et al. A microcircuitry comprised of microRNA-223 and transcription factors NF1-A and C/EBPα regulates human granulopoiesis. *Cell* **123**, 819–831 (2005).
- Johnnidis, J. B. et al. Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature* **451**, 1125–1129 (2008).
- Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
- John, B. et al. Human microRNA targets. *PLoS Biol.* **2**, e363 (2004).
- Betel, D., Wilson, M., Gabow, A., Marks, D. S. & Sander, C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* **36**, D149–D153 (2008).
- Lall, S. et al. A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.* **16**, 460–471 (2006).
- Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature Genet.* **39**, 1278–1284 (2007).
- Gaidatzis, D., van Nimwegen, E., Hausser, J. & Zavolan, M. Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics* **8**, 69 (2007).
- Miranda, K. C. et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203–1217 (2006).
- Robins, H., Li, Y. & Padgett, R. W. Incorporating structure to predict microRNA targets. *Proc. Natl Acad. Sci. USA* **102**, 4006–4009 (2005).
- Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. *Nature* **436**, 214–220 (2005).
- Long, D. et al. Potent effect of target structure on microRNA function. *Nature Struct. Mol. Biol.* **14**, 287–294 (2007).

32. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
33. Reddy, V. Y., Zhang, Q. Y. & Weiss, S. J. Pericellular mobilization of the tissue-destructive cysteine proteinases, cathepsins B, L, and S, by human monocyte-derived macrophages. *Proc. Natl Acad. Sci. USA* **92**, 3849–3853 (1995).
34. Menzel, K. *et al.* Cathepsins B, L and D in inflammatory bowel disease macrophages and potential therapeutic effects of cathepsin inhibition *in vivo*. *Clin. Exp. Immunol.* **146**, 169–180 (2006).
35. Fu, Y. K., Arkins, S., Wang, B. S. & Kelley, K. W. A novel role of growth hormone and insulin-like growth factor-I. Priming neutrophils for superoxide anion secretion. *J. Immunol.* **146**, 1602–1608 (1991).
36. Bjerknes, R. & Aarskog, D. Priming of human polymorphonuclear neutrophilic leukocytes by insulin-like growth factor I: increased phagocytic capacity, complement receptor expression, degranulation, and oxidative burst. *J. Clin. Endocrinol. Metab.* **80**, 1948–1955 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S.-J. Hong, T. Brummelkamp and S. Stehling-Sun for discussions, C. Bakalarski for writing and implementing the Vista algorithm for automated protein quantification, W. Johnston for technical assistance, and P. Wisniewski for cell sorting. This was supported by a Damon Runyon postdoctoral fellowship (C.S.) and grants from the NIH (D.P.B. and S.P.G.). D.P.B. is an investigator of the Howard Hughes Medical Institute.

Author Contributions The order of listing of the first three authors is arbitrary. C.S. and F.C. performed the experimental work with cells and animals. J.V. performed the mass spectrometry and associated computational analysis. D.B. performed the computational analysis of targeting. All authors contributed to the design of the study and preparation of the manuscript.

Author Information Array data and small RNA sequencing data were deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE12075. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.P.G. (steven_gygi@hms.harvard.edu) or D.P.B. (dbartel@wi.mit.edu).

METHODS

Transfection experiments. HeLa cells (ATCC, CCL-2) were grown in SILAC DMEM media (Invitrogen) supplemented with Pro (10 mg l^{-1}) and containing either naturally occurring isotopes of Arg and Lys (50 mg l^{-1} each) or heavy ($^{13}\text{C}_6$)-labelled Arg and Lys (50 mg l^{-1} each, Cambridge Isotope Laboratory). Heavy isotope incorporation in proteins was analysed by mass spectrometry ($>99\%$ Arg, $>98.5\%$ Lys). Cells grown in heavy amino acids were mock-transfected with lipofectamine 2000 (Invitrogen), whereas those grown in light amino acids were transfected with miRNA duplexes described previously^{7,9}, using lipofectamine 2000, 25 nM duplex, and supplementing with OPTI-MEM (Invitrogen). After 6 h, media of both mock and miRNA transfections was replaced with SILAC DMEM. Twenty-four hours after transfection, some cells were harvested, and mRNA was purified (RNeasy Plus, Qiagen) for expression profiling (Agilent human 4 × 44K microarray). Forty-eight hours after transfection, the remaining cells were harvested, and equal numbers of miRNA- and mock-transfected cells were mixed. Soluble nuclear proteins were purified (NEPER Nuclear and Cytoplasmic Extraction Reagent, Thermo Fisher Scientific) and separated into ten fractions by SDS-PAGE for mass spectrometry analysis.

As an additional control for targeting specificity, analyses of the transfection results were repeated comparing the response of proteins from messages with sites to the cognate miRNA to that of the very same proteins when the non-cognate miRNA was transfected. The overall conclusions from this set of control analyses were the same as for the analyses presented, indicating that the results depended on the identity of the miRNA transfected, rather than on other differences between mock- and miRNA-transfected cells, such as the mass of the amino acids or the presence of OPTI-MEM in the transfections.

Neutrophil culture and differentiation. All animal experiments were approved by the MIT Committee on Animal Care. Bone marrow was obtained from three 3-month-old wild-type male mice and from three 3-month-old *mir-223*^{-/-} mice²¹, and bone marrow haematopoietic progenitors were isolated as follows. Bone marrow from the three mice of each genotype was pooled, and suspended cells were depleted of mature cells using a mixture of biotin-conjugated Ter 119, Mac-1, Gr-1, B220 and CD3e antibodies (eBioscience) and anti-biotin microbeads (Miltenyi Biotech, Inc.), followed by magnetic cell sorting (MACS, Miltenyi Biotech, Inc.). The remaining cells were collected and cultured in SILAC IMDM media (Invitrogen) supplemented with Pro (10 mg l^{-1}) and containing G-CSF (100 ng ml^{-1} , PeproTech) and SCF (50 ng ml^{-1} , PeproTech). Media containing light Arg and Lys (50 mg l^{-1} each) was used for cells derived from wild-type mice, and heavy media containing $^{13}\text{C}_6$ -Arg and $^{13}\text{C}_6$ -Lys (50 mg l^{-1} each) was used for *mir-223* knockout cells. Media was replaced every two days, and after six days SCF was withdrawn to arrest proliferation and induce additional differentiation. Forty-two hours later, cells were harvested, and dead cells were removed (Dead cell removal kit, Miltenyi Biotech). Neutrophil maturity and viability were analysed by flow cytometry (FACSCalibur, BD Biosciences) after staining with PE-conjugated anti-mouse c-Kit antibody (eBioscience), APC-conjugated anti-mouse Gr-1 antibody (eBioscience), and propidium iodide (Supplementary Fig. 4a). The homogeneity of the cell population was also checked by microscopy after Wright-Giemsa stain of cytospun neutrophils (Supplementary Fig. 4b). Mass spectrometry confirmed nearly quantitative ($>99\%$) incorporation of heavy Arg and Lys in cells cultured from the *mir-223*-deficient mice.

A fraction of each cell population was used to purify mRNA (RNeasy Plus, Qiagen) for expression profiling (Affymetrix mouse 430 2.0 microarray, Fig. 2c). Equal numbers of cells from each population were mixed, and soluble nuclear and cytoplasmic protein preparations were fractionated by SDS-PAGE and analysed independently by LC-MS/MS. Additional biological replicates (each starting with bone marrow pooled from one to four additional mice) were prepared from both wild-type and knockout mice and used for mRNA expression profiling, RNA blotting and immunoblotting. Some of these additional replicates were sorted using FACSaria (BD Biosciences) with PE-conjugated c-Kit antibody and APC-conjugated Gr1 antibody to generate subpopulations for monitoring *mir-223* expression (Fig. 2b), for monitoring fates after additional culture (Supplementary Fig. 4c) and for mRNA profiling (Table 1 and Supplementary Data 4). For comparison, neutrophils directly isolated from wild-type and mutant mice, using biotin-conjugated Gr-1 antibody and MACS (each biological replicate pooling cells from three mice), were examined using expression profiling, RNA blotting and immunoblotting.

Mass spectrometry analysis. Protein ($50 \mu\text{g}$) was reduced (5 mM DTT in 50 mM ammonium bicarbonate, pH 8.2, at 56°C , 30 min) and alkylated (15 mM iodoacetamide, in 50 mM ammonium bicarbonate in the dark at room temperature, $20\text{--}22^\circ\text{C}$, 25 min), and then separated into 10 fractions (HeLa samples) or 16 fractions (neutrophils) by SDS-PAGE. Each fraction was in-gel digested with trypsin ($5 \text{ ng } \mu\text{l}^{-1}$ in 50 mM ammonium bicarbonate, pH 8.2, at 37°C , 16 h). Peptides were extracted in 50% acetonitrile (ACN) and 5% formic acid (FA), and

then dried down and desalted by reverse-phase (C18 StageTip). Peptide mixtures were resuspended in 5% ACN and 4% FA, and 20% of each mixture was analysed by LC-MS/MS in duplicate. Peptides were separated across a 55-min gradient ranging from 7% to 30% ACN in 0.1% FA in a microcapillary ($125 \mu\text{m} \times 17 \text{ cm}$) column packed with C18 reverse-phase material (Magic C18AQ, $5 \mu\text{m}$ particles, 200 \AA pore size, Michrom Bioresources) and on-line analysed on a hybrid linear ion trap Orbitrap (LTQOrbitrap, ThermoElectron) mass spectrometer. For each cycle, one full mass spectrometry scan acquired at high mass resolution (60,000 at 400 m/z , AGC target = 1×10^6 , maximum ion injection time = 1,000 ms) in the orbitrap analyser was followed by 10 MS/MS spectra on the linear ion trap (AGC target = 5×10^3 , maximum ion injection time = 120 ms) from the ten most abundant ions. Fragmented precursor ions were dynamically excluded from further selection for 35 s. Ions were also excluded if their charge was either <2 or unassigned.

Protein database searches and peptide quantification. MS/MS spectra were searched against the IPI protein sequence database using the Sequest algorithm. Peptide matches were filtered to $<1\%$ false-discovery rate using a target-decoy database strategy and using as filters mass deviation (in p.p.m.), Sequest Xcorr and dCn scores, and excluding sequences containing simultaneously heavy and light versions of Lys and Arg residues. Peptides were quantified using in-house Vista software^{37,38} by peak-area integration, and heavy/light peptide ratios were calculated. Among the set of independent measurements retained for each protein, the median of the heavy/light ratio was defined as the protein fold change (Supplementary Data 1–4). Quality cutoffs were as follows: all measurements were required to have a Vista confidence score ≥ 75 and a signal-to-noise ratio (S/N) ≥ 6.0 , where the S/N parameter was calculated as the sum of S/N_{Heavy} and S/N_{Light} . Measurements for proteins quantified with only one peptide were required to pass a more stringent S/N cutoff of 10.0. For proteins quantified with multiple peptides, independent measurements from a single peptide were not allowed to exceed half of the total number of independent measurements (by eliminating those measurements with lower S/N); this ensured that measurements for more than one peptide would influence the median.

To link the protein fold change to our reference cDNA set, the genomic coordinates of proteins from the IPI database³⁹ were used, requiring ≥ 50 nucleotide overlap between the genomic coordinates of the protein and a reference cDNA. To correct for the overall displacement of heavy and light populations (presumably caused by slightly unequal cell mixtures), we identified the subset of the proteins deriving from messages without 6–8mer seed-matched 3'-UTR sites, computed the difference in the median of heavy and light peaks, and offset all the fold-changes (including those from messages with sites) by this difference. This normalization caused our reported fold-change distribution of the proteins with no seed-matched sites to centre on zero.

cDNA data sets of non-redundant 5'-UTR, ORF or 3'-UTR sequences. We obtained human full-length cDNAs from RefSeq⁴⁰ and H-Invitational⁴¹ databases, and aligned them against the human genome³⁹ using BLAT⁴². Functional cDNAs were enriched as described previously⁴³, discarding those without introns as well as those with a low alignment quality, multiple high-scoring matches to the human genome, a premature stop codon or an incomplete coding sequence. If cDNAs had overlapping 3' UTRs, those obtained from the RefSeq database were chosen. If more than one cDNA remained, the cDNA with the longest 3' UTR was retained. The resulting set of non-redundant cDNAs was designated the 'reference cDNAs' (Supplementary Data 5). Multiple reference cDNAs for a single gene were allowed if the genomic coordinates of their 3' UTRs did not overlap with each other. However, when performing analysis of sites in ORFs or 5' UTRs, only a single cDNA was arbitrary chosen (from among the RefSeq cDNA, when present) to represent the gene, to prevent double counting the contribution from a single site. The same criteria were used to choose a unique reference cDNA to match each quantified protein. To search for miRNA seed-matched sites, the genomic sequence of the reference cDNA (with introns removed) was used instead of the cDNA sequence itself. The analogous procedure was repeated for mouse full-length cDNAs, from RefSeq and FANTOM DB⁴⁴ databases, aligned against the mouse genome (Supplementary Data 5).

Microarray data processing. The 60-nucleotide probe sequences of Agilent 4 × 44K microarray were aligned against the human genome using BLAT. Any probe that had a less than a perfect match to the human genome or multiple perfect matches was removed. The mRNA fold change and the corresponding error, generated by the Agilent Feature Extraction Software, were linked to our reference cDNA set by a method analogous to that used for the SILAC data described previously (Supplementary Data 1–3). Similarly, a set of probe 'consensus sequences' from the Affymetrix mouse 430 2.0 microarray were aligned against the mouse genome. Any probe consensus sequence that had a BLAT alignment score of <100 or that had multiple high-scoring matches to the genome (that is, whose top two alignments to the genome had $<1\%$ difference in percentage identity) was removed. For each probe consensus sequence, the

mRNA fold change between the wild type and *mir-223*^{-/-} and its standard error were computed after quantile-normalizing the expression data from the multiple chips using the RMAExpress software⁴⁵ (Supplementary Data 4). When mapping the Agilent probes and Affymetrix probe consensus sequences to our reference cDNA set, ≥ 15 nucleotides of genomic coordinates between the probe and a reference cDNA were required to overlap.

Considerations for target site analyses. The minimal fraction of genes responding to the miRNA was calculated from cumulative distributions, determining the maximal cumulative difference between distributions, with correction for distribution bumpiness, as described⁷. To prevent undue impact from a few outliers, fold changes were truncated at ± 2.0 before calculating mean log-fold changes. To evaluate sequence conservation of human reference cDNAs, human, mouse, rat and dog alignments were extracted from 28 vertebrate genome alignments (aligned against the human genome) obtained from the UCSC Genome Bioinformatics Site⁴⁶. A site was considered conserved if also found in the orthologous positions of the other three genomes, allowing for horizontal shifts of the site (resulting from presumed artefacts or ambiguities in the alignment), provided that two of the alignment columns (each column being the width of one position in the alignment) overlapped the site in all four species. Similarly, from 30 vertebrate genome alignments (aligned against the mouse genome), the four mammalian sequences were extracted to assess the sequence conservation of mouse reference cDNAs and to identify conserved target sites.

Comparison of target prediction tools. Lists of miRNA targets predicted by TargetScan (release 4.1)^{2,7}, PicTar (human, chimp, mouse, rat, dog)^{4,25}, miRanda (January 2008 release)^{23,24}, miRBase Targets (version 5)²², RNA22 (ref. 28) and PITA²⁶ were obtained from their respective websites, using the most recent predictions publicly available as of March 2008. Most of these consisted of gene symbols, sequence identification of full-length cDNAs, and/or scores. To map these predictions to the human or mouse genome, genomic alignments of RefSeq, Ensembl and UCSC genes were obtained from the UCSC Genome Bioinformatics Site⁴⁶, and the most informative set of alignments for each prediction tool was used. To prevent double counting, a single prediction was arbitrarily chosen for genes with multiple redundant predictions.

Small-RNA sequencing, RNA blots and protein blots. Small RNAs were sequenced on the Solexa platform using a protocol modified from that used previously⁴⁷. RNA blots analysed 5 μ g total RNA per lane and used carbodiimide-mediated cross-linking to the membrane⁴⁸. Protein blots were probed using the following antibody dilutions: anti-CstII goat monoclonal antibody (R&D Systems), 1:1,600; anti-Igflr rabbit polyclonal antibody (Santa Cruz Biotechnology), 1:1,000; anti-Cbx5 (HP-1 α) mouse monoclonal antibody (Millipore), 1:2,500; anti-actin mouse monoclonal antibody (Abcam), 1:25,000; and anti-actin rabbit polyclonal antibody (Cell Signaling), 1:10,000.

37. Everley, P. A. *et al.* Enhanced analysis of metastatic prostate cancer using stable isotopes and high mass accuracy instrumentation. *J. Proteome Res.* **5**, 1224–1231 (2006).
38. Guo, A. *et al.* Signaling networks assembled by oncogenic EGFR and c-Met. *Proc. Natl Acad. Sci. USA* **105**, 692–697 (2008).
39. Kersey, P. J. *et al.* The International Protein Index: an integrated database for proteomics experiments. *Proteomics* **4**, 1985–1988 (2004).
40. Pruitt, K. D., Katz, K. S., Sicotte, H. & Maglott, D. R. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. *Trends Genet.* **16**, 44–47 (2000).
41. Imanishi, T. *et al.* Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.* **2**, e162 (2004).
42. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
43. Baek, D., Davis, C., Ewing, B., Gordon, D. & Green, P. Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res.* **17**, 145–155 (2007).
44. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
45. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
46. Karolchik, D. *et al.* The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* **36**, D773–D779 (2008).
47. Ruby, J. G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**, 1193–1207 (2006).
48. Pall, G. S., Codony-Servat, C., Byrne, J., Ritchie, L. & Hamilton, A. Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.* **35**, e60 (2007).

ARTICLES

Type IV collagens regulate BMP signalling in *Drosophila*

Xiaomeng Wang¹, Robin E. Harris¹, Laura J. Bayston¹ & Hilary L. Ashe¹

Dorsal–ventral patterning in vertebrate and invertebrate embryos is mediated by a conserved system of secreted proteins that establishes a bone morphogenetic protein (BMP) gradient. Although the *Drosophila* embryonic Decapentaplegic (Dpp) gradient has served as a model to understand how morphogen gradients are established, no role for the extracellular matrix has been previously described. Here we show that type IV collagen extracellular matrix proteins bind Dpp and regulate its signalling in both the *Drosophila* embryo and ovary. We provide evidence that the interaction between Dpp and type IV collagen augments Dpp signalling in the embryo by promoting gradient formation, yet it restricts the signalling range in the ovary through sequestration of the Dpp ligand. Together, these results identify a critical function of type IV collagens in modulating Dpp in the extracellular space during *Drosophila* development. On the basis of our findings that human type IV collagen binds BMP4, we predict that this role of type IV collagens will be conserved.

BMPs are secreted signalling molecules that belong to the TGF- β superfamily of growth factors. BMPs are used repeatedly throughout development of many invertebrates and vertebrates in a variety of biological processes, including the patterning of different tissues and organs^{1,2}. As a result, altered BMP signalling has been associated with several human diseases, including skeletal disorders, vascular diseases and cancer^{3–5}. Moreover, the therapeutic potential of BMPs in bone and cartilage repair is being explored at present^{6,7}.

In *Drosophila*, the BMP signalling molecule Dpp specifies cell fates at different developmental stages⁸. In the *Drosophila* embryo, a heterodimer of the BMP signalling molecules Dpp and Screw (Scw) patterns the dorsal ectoderm into different cell types. Although uniformly expressed in the dorsal ectoderm, Dpp/Scw heterodimers are redistributed to generate a concentration gradient⁹. Dpp signalling is also essential for the maintenance of germline stem cells (GSCs) in the germarium of the *Drosophila* ovary. Cells in the stem cell niche provide a short-range Dpp signal to inhibit differentiation¹⁰.

Studies in the *Drosophila* embryo have suggested that Dpp diffusion is restricted in the absence of a protein shuttling complex^{11–13}. One possibility to explain this could be that Dpp is immobilized by an extracellular matrix protein. Therefore, we evaluated potential roles of extracellular matrix proteins in regulating Dpp signalling. By combining genetic, biochemical and transgenic approaches, we have obtained evidence that type IV collagen, the main component of a specialized form of extracellular matrix¹⁴, is critical for correct Dpp signalling during *Drosophila* development.

Type IV collagen proteins bind Dpp

There are two type IV collagen proteins in *Drosophila*, Viking (Vkg)¹⁵ and Dcg1 (also known as Cg25C; ref. 16). We tested the ability of secreted biologically active, epitope-tagged Dpp purified from the media of transfected *Drosophila* S2 tissue culture cells¹⁷ to bind to the amino- and carboxy-terminal non-collagenous domains of Vkg and Dcg1 (Fig. 1a). Dpp–haemagglutinin (HA) binds to the C-terminal but not the N-terminal domains of both Vkg and Dcg1 (Fig. 1a), whereas denatured Dpp–HA protein does not (Supplementary Fig. 1a). Dpp/Scw heterodimers also bind to the Vkg and Dcg1 C-terminal domains (Fig. 1b; VkgC and DcgC).

Surface plasmon resonance analyses show that the binding between Dpp and glutathione *S*-transferase (GST)–VkgC or GST–DcgC is saturable and has dissociation constants (K_d) of 0.75 and 0.65 μ M, respectively (Supplementary Fig. 2).

Deletion analysis of VkgC identified a region required for Dpp interaction (Supplementary Fig. 1b) which, when aligned with the equivalent region of Dcg1, shows a short conserved sequence (Fig. 1c). Deletion of five of these amino acids from the Vkg C-terminal domain severely attenuates the interaction between Vkg and both Dpp and Dpp/Scw ligands (Fig. 1c). As these binding studies used GST fusion proteins purified from bacteria, we have confirmed the results using GST–VkgC and GST–VkgC Δ proteins secreted into the media of transfected S2 cells (Supplementary Fig. 3). In addition, further mutational analysis of the Vkg sequences required for Dpp interaction is shown in Supplementary Fig. 4. The amino acids in Vkg that are necessary for Dpp interaction are present in a sequence that is conserved in mosquito, worm, mouse and human type IV collagens (Fig. 1d). Alignment of all the known type IV collagen sequences from these species identifies a consensus Y/FI/VSRCXVCE, which may function as a conserved BMP-binding module. In support of this, we have shown saturable binding between human full-length triple-helical type IV collagen and BMP4 with a K_d of 92 nM (Supplementary Fig. 2).

Type IV collagen distribution

We investigated the expression and distribution of collagen IV during two developmental stages when Dpp signalling occurs. Using semiquantitative PCR with reverse transcription (RT–PCR), we obtained evidence for maternal, but not early zygotic, transcription of *vkg* and *Dcg1* (Supplementary Fig. 5a). Maternal expression of type IV collagens has been reported previously^{18,19}. Translation of the *vkg* and *Dcg1* transcripts would result in deposition of type IV collagens in the early embryo. To visualize this, we analysed embryos carrying a green fluorescent protein (GFP) exon in the endogenous *vkg* gene²⁰ using an anti-GFP antibody. As a control, PatJ protein—which has a basal localization in the epithelium of early embryos²¹—was also detected. Specific Vkg–GFP staining can be detected in transgenic, but not control, early (Supplementary Fig. 5b) and gastrulating

¹Faculty of Life Sciences, The University of Manchester, Oxford Road, Manchester M13 9PT, UK.

differentiating cytotoblasts and cysts (Fig. 2d). Analysis of the Vkg–GFP distribution demonstrates GFP staining around all the somatic cells of Vkg–GFP, but not control, germaria (Fig. 2e). Within each germarium, Vkg–GFP is also detected between somatic niche cells and the GSCs. Here, the GSCs are recognized by their position and size, as well as the presence and position of their unique spectrosome organelle (data not shown).

Collagen IV augments embryonic Dpp signalling

We next used these two developmental contexts to assess the impact of type IV collagens on Dpp signalling. First we focused on the embryo and analysed the expression patterns of different Dpp threshold responses²³ in embryos from *vkg*^{k00236} or *Dcg1*^{k00405} heterozygous females, as both type IV collagen genes are maternally expressed. Expression of the peak Dpp thresholds, *Race* (also known as *Ance*) and *hindsight* (*hnt*; also known as *peb*), is lost in the presumptive amnioserosa, whereas the *Race* head spots and posterior *hnt* expression persist, as these are less responsive to changes in Dpp signalling²³ (Fig. 3a). The expression pattern of the lower threshold response *u-shaped* (*ush*) is thinner than the wild-type pattern. These phenotypes resemble those of embryos heterozygous for the null *dpp*^{Hin37} allele²³ or homozygous for the hypomorphic *dpp*^{hr56} allele (Supplementary Fig. 6), consistent with a lower amount of Dpp signalling in the mutant embryos.

This phenotype is due to a maternal effect caused by a lower *vkg* or *Dcg1* dose from the heterozygous females, as no effect on Dpp target gene expression patterns is observed in embryos from the reciprocal crosses (Fig. 3a). The expression patterns of other Dpp targets, including *zerknüllt* and *tailup*, are also affected in the mutant embryos (Supplementary Fig. 7). Similar effects on Dpp target gene expression patterns were observed in mutant embryos when three different *vkg* alleles were tested (data not shown). It was not possible to analyse embryos with a further reduction in Vkg or *Dcg1*, as females *trans*-heterozygous for different *vkg* alleles, or homozygous for the weak *vkg*¹⁹⁷ mutation, are sterile. Further support for the link between type IV collagens and Dpp signalling is provided by our *vkg* and *dpp* genetic interaction data (Supplementary Fig. 8). Moreover, *Race* expression in the presumptive amnioserosa of embryos from *vkg*/+ females can be rescued by *dpp* overexpression (Supplementary Fig. 9), further corroborating the evidence presented thus far that Dpp signalling is compromised in the mutant embryos.

To investigate the defect in Dpp signalling further, we directly visualized Dpp, using a *dpp*–HA transgene¹², and its activated signal transducer phosphorylated Mad (pMad) in mutant embryos. Both the Dpp and pMad protein domains are narrower in embryos from *vkg* heterozygous females than those from wild-type females (Fig. 3b). We also detected Short gastrulation (Sog) protein, as Sog has an important role in Dpp/Scw gradient formation²⁴. However, Sog distribution is unaffected in mutant embryos (Fig. 3b). Overall, the data are consistent with type IV collagens augmenting Dpp signalling in the embryo, through an interaction between type IV collagens and Dpp.

Collagen IV restricts Dpp signalling range in the ovary

Because we also detected a Vkg extracellular matrix around the ovary GSCs that are maintained by Dpp signalling¹⁰, we analysed the number of GSCs and cytotoblasts in germaria from females that are wild type, *trans*-heterozygous for different *vkg* mutant alleles or homozygous for the *vkg*¹⁹⁷ allele. Analysis of *vkg*^{k07138}/*vkg*⁰¹²⁰⁹ mutant germaria, for example, shows an increased number of GSCs, as visualized by the presence of a spectrosome and the absence of Bam protein (Fig. 4a). Quantification demonstrates that there is an increase in GSC number in all mutant germaria analysed (Fig. 4b). As GSC number is increased by overexpression of Dpp²⁵, our data are consistent with type IV collagens being required for correct Dpp signalling in the germarium. We propose that in mutant germaria with reduced Vkg levels Dpp is not sequestered around the GSCs, resulting in an extended signalling range that causes ectopic GSC fate (see Discussion).

Mechanism of regulation of embryonic Dpp signalling

Our data from the early embryo clearly suggest that type IV collagens increase Dpp signalling (Fig. 3). In contrast to the ovary germarium, in which there is a localized source of Dpp, in the embryo a Dpp/Scw gradient is formed across a field of cells in which these two ligands are uniformly expressed⁹. Gradient formation is dependent on two BMP-binding proteins, Sog²⁴ and Twisted gastrulation (Tsg)¹⁷. As any model invoking an interaction between secreted Dpp and type IV collagens needs to also factor in the functions of Sog and Tsg, we tested the ability of Sog or Tsg to bind to Vkg. Like Dpp, Sog interacts with the C terminus of Vkg, whereas Tsg does not (Fig. 5a). To understand the dynamics of Sog and Dpp/Scw interactions with Vkg, Dpp/Scw heterodimers were bound to the GST–VkgC fusion protein immobilized on beads, and the complex was washed to remove any unbound Dpp/Scw. This was followed by the addition of either Sog or Tsg, both Sog and Tsg, or negative control medium from mock-transfected cells (Fig. 5b); the amount of Dpp present in both the supernatant and bead fractions was then determined. The addition of control medium or of Sog or Tsg alone had no effect on the Dpp/Scw heterodimer, which remained bound to the GST–Vkg beads (Fig. 5b). When Sog alone was added, it was also able to bind to the Vkg beads. However, when Sog and Tsg were added together, they mediated the release of Dpp/Scw heterodimers from Vkg (Fig. 5b). Formation of such a Dpp/Scw–Sog–Tsg complex is an essential step in Dpp/Scw gradient formation⁹.

We also investigated whether type IV collagens affect Dpp–receptor interactions. For this, we used a commercially available mouse BMP receptor 1A (BMPR) protein that has been used previously to investigate receptor binding²⁶. This BMPR overcomes the difficulties associated with purifying functional receptors, and mouse BMP receptors can function in the Dpp pathway²⁷. Binding of Dpp to the BMPR protein in the absence and presence of increasing amounts of the VkgC protein fragment was investigated by immunoprecipitation

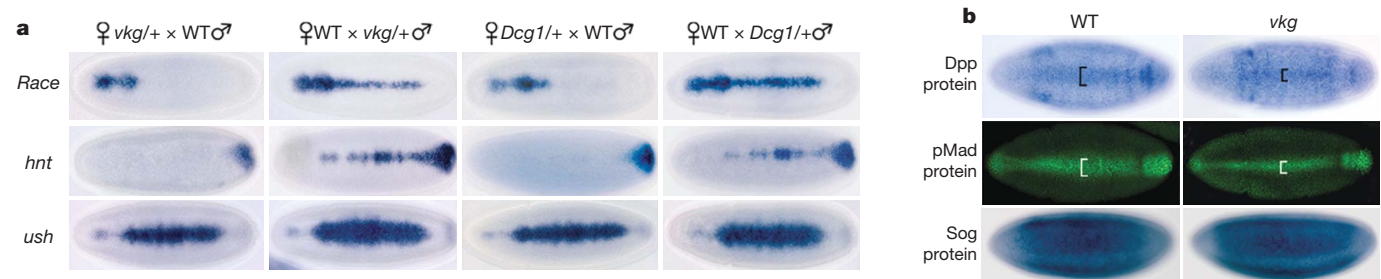


Figure 3 | Type IV collagens increase Dpp signalling in the embryo. **a**, RNA *in situ* hybridization of embryos from *vkg*^{k00236} or *Dcg1*^{k00405} heterozygous females crossed to wild-type (WT) males, or the reciprocal crosses, as labelled. Embryos (dorsal views) are at the onset of gastrulation and show *Race*, *hnt* or *ush* expression. **b**, Immunostaining of Dpp (using a Dpp–HA

transgene), pMad and Sog distribution in embryos from *vkg*^{k00236}/+ females (see Methods for details of Dpp–HA embryos). Embryos are shown as dorsal (Dpp and pMad) and lateral (Sog) views. The extent of the Dpp and pMad stripes is marked by brackets. Original magnification, ×20.

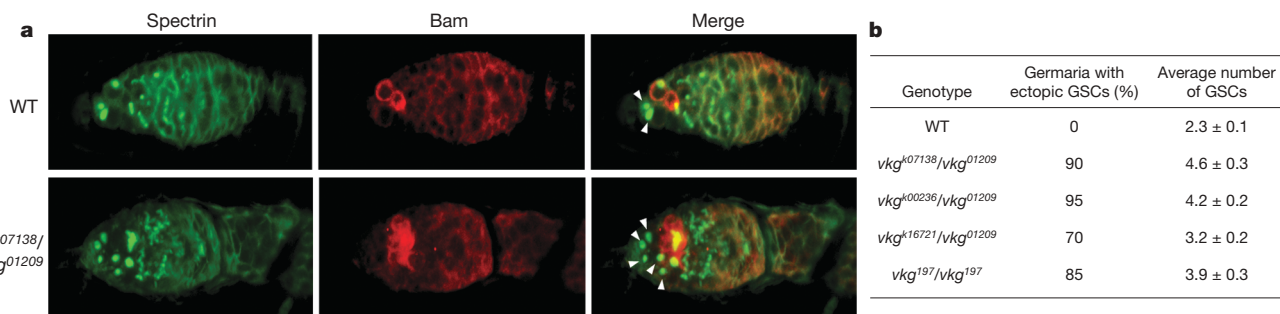


Figure 4 | Collagen IV increases GSC number. **a**, Confocal images of wild-type and *vkg^{k07138}/vkg⁰¹²⁰⁹* mutant ovaries showing Spectrin and Bam staining. The GSCs are labelled with arrowheads. Original magnification, ×63. **b**, Table showing the proportion of mutant ovaries observed from the

various females tested, along with the average number of GSCs found in the mutant ovaries ($n = 20$) ± s.e.m. In all mutant ovaries, two cystoblasts are typically observed as in wild type.

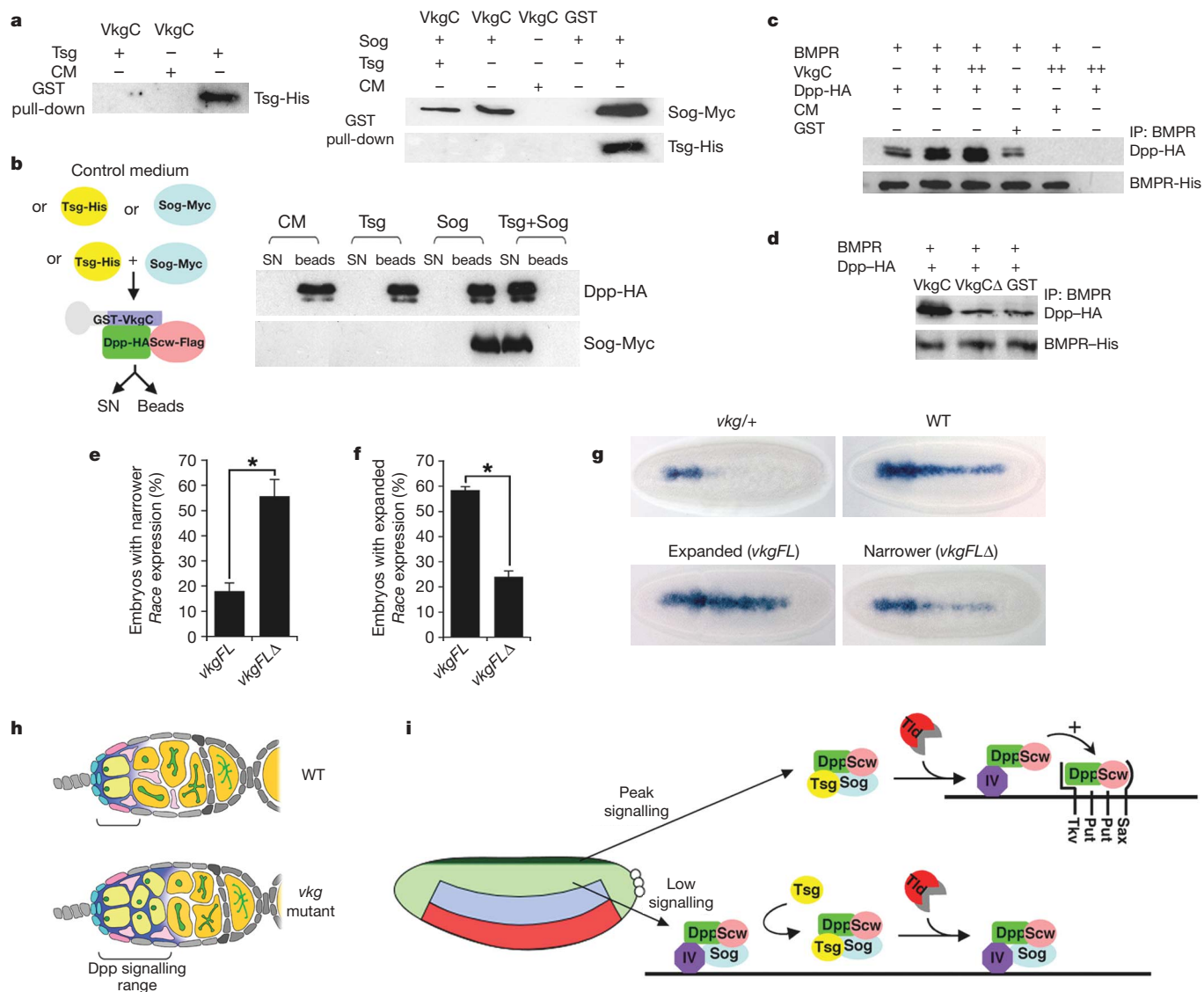


Figure 5 | Mechanistic insight into the function of Dpp–collagen-IV interactions in the embryo. **a**, Western blot analysis of interactions between Tsg–His (left) or Sog–Myc (right) and VkgC. **b**, Schematic of the experimental scheme (left), with western blots (right) showing Dpp–HA and Sog–Myc in the supernatant (SN) and bead fractions. **c**, Western blot showing Dpp–HA immunoprecipitated (IP) by the BMP receptor (BMPR) in the presence of increasing amounts of GST–VkgC or control GST. Control medium (CM) instead of Dpp, and a reaction lacking the receptor, are included as negative controls. The immunoprecipitated receptor levels are also shown. **d**, As in **c**, except that GST–VkgCΔ was tested. **e**, **f**, Graphs

showing the percentage of embryos from *vkg/+*; *vkgFL* and *vkg/+*; *vkgFLΔ* females that have a narrower (**e**) or expanded (**f**) *Race* expression pattern. Each bar is the average of data obtained from counts of at least 40 embryos from each of three independent transgenic lines ($n = 3$). Error bars denote s.e.m.; asterisk, $P < 0.05$, Student's *t*-test. **g**, Representative embryos with narrower or expanded *Race* expression patterns as counted in **e** and **f**, with control wild-type embryos or those from *vkg/+* females for comparison. Original magnification, ×20. **h**, **i**, Models of the regulation of Dpp signalling by type IV collagens (IV) in the ovary and embryo, respectively.

and western blot analysis. The data show that more Dpp binds to this type I receptor in the presence of increasing Vkg (Fig. 5c). We infer from this observation that a second function of type IV collagens is to promote the interaction of Dpp/Scw with its heteromeric receptor complex. Furthermore, the VkgCΔ protein, which has greatly reduced capacity to bind Dpp (Fig. 1c), does not promote receptor binding (Fig. 5d).

Next we tested the effect of overexpressing full-length Vkg in wild-type embryos from a transgene containing the *ubiquitin* enhancer to direct maternal expression²⁸, similar to that of endogenous *vkg*. *vkg* overexpression in wild-type embryos led to an expansion in *Race* expression (Supplementary Fig. 10a), consistent with enhanced gradient formation and/or receptor interactions. As a critical test of the *in vivo* requirement for the sequences in Vkg that we have identified as being important for Dpp/Scw binding *in vitro* (Fig. 1c), we generated a mutant *vkg* transgene that lacks five of the conserved amino acids required for the Dpp–Vkg interaction. The wild-type (*vkgFL*) and mutant (*vkgFLΔ*) transgenes were crossed into *vkg/+* females, and the effect on *Race* expression in embryos was analysed. Three different classes of *Race* expression patterns—narrower, wild type or expanded—were observed and counted in embryos from the *vkg/+*; *vkgFL(Δ)* females.

The proportion of *Race*-expressing embryos classified as narrower was expressed as a percentage of the proportion of mutant embryos from *vkg/+* females that lack a transgene (see Methods). The data are shown in Fig. 5e, in which each bar represents an average from three independent transgenic lines. There are fewer embryos with narrower *Race* expression patterns when *vkgFL* is expressed compared to the *vkgFLΔ* transgene (Fig. 5e). An example of the type of *Race* expression pattern scored as narrower is shown in Fig. 5g. The low percentage of embryos remaining with narrower *Race* expression when the wild-type *vkgFL* transgene is present (Fig. 5e) probably reflects a difference in the expression level from the *ubiquitin* enhancer.

The proportions of embryos from *vkg/+*; *vkgFL(Δ)* females with expanded *Race* expression patterns (see Fig. 5g for an example) were quantified and expressed relative to the typical proportion of embryos from *vkg/+* females with wild-type *Race* expression patterns (Fig. 5f). The wild-type transgene led to a significantly higher proportion of embryos with expanded *Race* expression than the mutant transgene. The residual activity of the mutant transgene may in part reflect the low capacity of the VkgCΔ protein to interact with Dpp/Scw *in vitro*, albeit at a greatly reduced amount relative to VkgC (Fig. 1c). In addition, both transgenes are expressed in embryos with wild-type levels of Dcg1 and in which some endogenous Vkg protein remains. The C-terminal domains of type IV collagens interact to mediate heterotrimer formation and assembly of the heterotrimers into a network¹⁴. Therefore, the relative stoichiometry of associated wild-type Dcg1/Vkg and mutant Vkg C-terminal domains which Dpp/Scw encounters may ultimately influence binding *in vivo*. Nevertheless, the clear difference in the behaviour of the wild-type and mutant Vkg proteins provides *in vivo* support for the *in vitro* binding data. Moreover, these transgenic data suggest that the important role for type IV collagens in the regulation of Dpp/Scw signalling in the embryo is mediated by a direct interaction. Further support for an important Dpp–Vkg interaction comes from the different capacities of the two transgenes to rescue *dpp^{hr56}* homozygous embryos (Supplementary Fig. 10b).

Discussion

Our data demonstrate that the interaction between Dpp and collagen IV is an essential aspect of correct signalling in the *Drosophila* gerarium and early embryo. In wild-type geraria, we suggest that Dpp secreted from the niche binds to Vkg, which restricts Dpp signalling range from the source. In mutant geraria with reduced Vkg protein, less Dpp will be bound by Vkg, resulting in an increased Dpp signalling range which downregulates *bam* transcription in more cells, thereby increasing GSC number (Fig. 5h).

In the embryo, we favour a model whereby binding of Dpp/Scw to type IV collagens facilitates assembly of the Dpp/Scw–Sog–Tsg complex. Tollid (Tld) cleavage of this complex releases Dpp/Scw, which can rebind type IV collagens. In the presence of Sog, the inhibitory complex will be reassembled, whereas in the absence of Sog, type IV collagens will promote Dpp/Scw–receptor interactions (Fig. 5i). This latter function may require the unusual apical distribution of Vkg protein in the embryo, as Dpp seems to predominantly interact with its receptor apically in the embryo¹³. In the dorsal ectoderm, initial Dpp signalling enhances subsequent Dpp signalling by the activation of an as yet unidentified target gene in a positive feedback loop¹³. By promoting Dpp–receptor interactions at the dorsal midline leading to target gene activation, type IV collagens will facilitate the amplification of signalling by positive feedback.

The model explains the phenotype of embryos from *vkg/+* females, as the reduced amount of type IV collagens would impair assembly of the Dpp/Scw–Sog–Tsg complex and initial gradient formation. Disruption of the early gradient, in combination with reduced receptor interactions in type IV collagen mutant embryos, will reduce target gene expression and positive feedback, further decreasing subsequent signalling. As a result, the peak Dpp target genes are lost and intermediate thresholds are thinner.

In addition to the role of type IV collagens in regulating Dpp signalling in the early embryo that we have described here, integrins—another principal constituent of basal lamina—are required for apposition of the amnioserosa and yolk sac to mediate proper germ band retraction and dorsal closure during later embryonic development^{29,30}. Therefore, basal lamina components have repeated roles in dorsal–ventral patterning of the fly embryo. Different types of extracellular matrix proteins also modulate BMP signalling at other development stages, for example, heparan sulphate proteoglycans regulate Dpp movement in the *Drosophila* wing³¹. In vertebrates, type IV collagens are not only transcriptional targets of BMP signalling³², but they also bind BMP4 (Supplementary Fig. 2) and have been suggested to potentiate signalling in tissue culture cells³³. We propose that the conserved sequence that we have identified will function as a BMP-binding module, and that type IV collagens will affect BMP signalling during vertebrate development.

METHODS SUMMARY

Drosophila stocks. The fly stocks used were: *vkg^{k00236}*, *vkg^{k16721}*, *vkg^{k07138}*, *vkg⁰¹²⁰⁹*, *vkg¹⁹⁷*, *vkg^{G454}*, *Dcg1^{k00405}*, *dpp^{hr56}* and genomic Dpp–HA (ref. 12). The severity of the alleles tested in terms of the effect on Dpp targets in the embryo is *vkg^{k00236}* > *vkg^{k1672}* > *vkg¹⁹⁷* = *vkg⁰¹²⁰⁹* (data not shown). *yw⁶⁷²³* flies were used as wild type. The *vkgFL* and *vkgFLΔ* transgenic lines were generated using standard methods.

RNA *in situ* hybridization and immunostaining. Embryo collection, ovary dissection, RNA *in situ* hybridizations using digoxigenin-labelled RNA probes and immunostaining were performed using standard methods. Antibodies used for immunostaining and the strategy for calculating the expression pattern data graphed in Fig. 5e, f are described in Methods.

Plasmids, transfection, protein purification and GST pull-downs. Plasmids generated in this study are described in Methods. The Dpp–HA, Sog–Myc, Tsg–His and Scw–Flag plasmids have been described^{12,34} (provided by M. O'Connor). S2 cells were transfected using Effectene transfection reagent (Qiagen) and the epitope-tagged proteins were purified as described¹². GST fusion proteins were purified from *Escherichia coli* BL21 transformed with the appropriate pGEX–Vkg/Dcg1 plasmid according to the manufacturer's protocol (GE Healthcare). The pull-down method is described in Supplementary Methods. The BMPR immunoprecipitation was performed as described²⁶, except that the BMPR was bound to protein A sepharose at the start of the assay, and proteins were eluted at 4 °C for 2 h.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 February; accepted 26 June 2008.

Published online 13 August 2008.

- Hogan, B. L. Bone morphogenetic proteins in development. *Curr. Opin. Genet. Dev.* 6, 432–438 (1996).

2. Kishigami, S. & Mishina, Y. BMP signaling and early embryonic patterning. *Cytokine Growth Factor Rev.* **16**, 265–278 (2005).
3. Chen, D., Zhao, M. & Mundy, G. R. Bone morphogenetic proteins. *Growth Factors* **22**, 233–241 (2004).
4. Bobik, A. Transforming growth factor- β s and vascular disorders. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1712–1720 (2006).
5. Harradine, K. A. & Akhurst, R. J. Mutations of TGF β signaling molecules in human disease. *Ann. Med.* **38**, 403–414 (2006).
6. Reddi, A. H. BMPs: from bone morphogenetic proteins to body morphogenetic proteins. *Cytokine Growth Factor Rev.* **16**, 249–250 (2005).
7. Bessa, P. C., Casal, M. & Reis, R. L. Bone morphogenetic proteins in tissue engineering: the road from laboratory to clinic, part II (BMP delivery). *J. Tissue Eng. Regen. Med.* **2**, 81–96 (2008).
8. Podos, S. D. & Ferguson, E. L. Morphogen gradients: new insights from DPP. *Trends Genet.* **15**, 396–402 (1999).
9. Ashe, H. L. BMP signalling: synergy and feedback create a step gradient. *Curr. Biol.* **15**, R375–R377 (2005).
10. Kirilly, D. & Xie, T. The *Drosophila* ovary: an active stem cell community. *Cell Res.* **17**, 15–25 (2007).
11. Eldar, A. *et al.* Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning. *Nature* **419**, 304–308 (2002).
12. Shimmi, O., Umulis, D., Othmer, H. & O'Connor, M. B. Facilitated transport of a Dpp/Scw heterodimer by Sog/Tsg leads to robust patterning of the *Drosophila* blastoderm embryo. *Cell* **120**, 873–886 (2005).
13. Wang, Y. C. & Ferguson, E. L. Spatial bistability of Dpp-receptor interactions during *Drosophila* dorsal–ventral patterning. *Nature* **434**, 229–234 (2005).
14. Khoshnoodi, J., Cartiailler, J. P., Alvares, K., Veis, A. & Hudson, B. G. Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers. *J. Biol. Chem.* **281**, 38117–38121 (2006).
15. Yasothornsrikul, S., Davis, W. J., Cramer, G., Kimbrell, D. A. & Dearolf, C. R. *viking*: identification and characterization of a second type IV collagen in *Drosophila*. *Gene* **198**, 17–25 (1997).
16. Blumberg, B., MacKrell, A. J. & Fessler, J. H. *Drosophila* basement membrane procollagen α 1(IV). Complete cDNA sequence, genomic structure, and general implications for supramolecular assemblies. *J. Biol. Chem.* **263**, 18328–18337 (1988).
17. Ross, J. J. *et al.* Twisted gastrulation is a conserved extracellular BMP antagonist. *Nature* **410**, 479–483 (2001).
18. Knibiehler, B., Mirre, C. & Le Parco, Y. Collagen type IV of *Drosophila* is stockpiled in the growing oocyte and differentially located during early stages of embryogenesis. *Cell Differ. Dev.* **30**, 147–157 (1990).
19. Medioni, C. & Noselli, S. Dynamics of the basement membrane in invasive epithelial clusters in *Drosophila*. *Development* **132**, 3069–3077 (2005).
20. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 15050–15055 (2001).
21. Bhat, M. A. *et al.* Discs Lost, a novel multi-PDZ domain protein, establishes and maintains epithelial polarity. *Cell* **96**, 833–845 (1999).
22. Lunstrum, G. P. *et al.* *Drosophila* basement membrane procollagen IV. *J. Biol. Chem.* **263**, 18318–18327 (1988).
23. Ashe, H. L., Mannervik, M. & Levine, M. Dpp signaling thresholds in the dorsal ectoderm of the *Drosophila* embryo. *Development* **127**, 3305–3312 (2000).
24. Ashe, H. L. & Levine, M. Local inhibition and long-range enhancement of Dpp signal transduction by Sog. *Nature* **398**, 427–431 (1999).
25. Xie, T. & Spradling, A. C. *decapentaplegic* is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. *Cell* **94**, 251–260 (1998).
26. Larraín, J. *et al.* BMP-binding modules in chordin: a model for signalling regulation in the extracellular space. *Development* **127**, 821–830 (2000).
27. Maduzia, L. L. & Padgett, R. W. *Drosophila* MAD, a member of the Smad family, translocates to the nucleus upon stimulation of the dpp pathway. *Biochem. Biophys. Res. Commun.* **238**, 595–598 (1997).
28. Lee, H. S., Simon, J. A. & Lis, J. T. Structure and expression of ubiquitin genes of *Drosophila melanogaster*. *Mol. Cell. Biol.* **8**, 4727–4735 (1988).
29. Schock, F. & Perrimon, N. Retraction of the *Drosophila* germ band requires cell–matrix interaction. *Genes Dev.* **17**, 597–602 (2003).
30. Reed, B. H., Wilk, R., Schock, F. & Lipshitz, H. D. Integrin-dependent apposition of *Drosophila* extraembryonic membranes promotes morphogenesis and prevents anoikis. *Curr. Biol.* **14**, 372–380 (2004).
31. Belenkaya, T. Y. *et al.* *Drosophila* Dpp morphogen movement is independent of dynamin-mediated endocytosis but regulated by the glypican members of heparan sulfate proteoglycans. *Cell* **119**, 231–244 (2004).
32. Abe, H. *et al.* Type IV collagen is transcriptionally regulated by Smad1 under advanced glycation end product (AGE) stimulation. *J. Biol. Chem.* **279**, 14201–14206 (2004).
33. Paralkar, V. M., Weeks, B. S., Yu, Y. M., Kleinman, H. K. & Reddi, A. H. Recombinant human bone morphogenetic protein 2B stimulates PC12 cell differentiation: potentiation and binding to type IV collagen. *J. Cell Biol.* **119**, 1721–1728 (1992).
34. Yu, K. *et al.* Processing of the *Drosophila* Sog protein creates a novel BMP inhibitory activity. *Development* **127**, 2143–2154 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. O'Connor for plasmids, flies and advice on Dpp–HA staining; W. Chia, R. Ray and the Bloomington stock centre for fly stocks; E. Bier, P. ten Dijke, H. Bellen, L. Fessler and D. McKearin for antibodies; T. Jowitt for help with SPR; S. Lunj, M. Ronshaugen and W. Song for help; M. Ashe and W. Miles for reading the manuscript; and M. Levine whose laboratory this work was initiated in. This work was funded by a BBSRC project grant to H.L.A. and a BBSRC studentship to R.E.H.

Author Contributions H.L.A. designed and obtained funding for the experiments. X.W. generated the *in vitro* and embryonic data. R.E.H. performed the ovary experiments. L.J.B. provided technical assistance with cloning, fly counts and transgenics. X.W., R.E.H. and H.L.A. analysed and interpreted the data, and H.L.A. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.L.A. (hilary.ashe@manchester.ac.uk).

METHODS

Plasmid construction. To generate plasmids for expression of GST-Vkg or GST-Dcg1 fusion proteins, various *vkg* or *Dcg1* complementary DNA fragments were isolated by PCR and cloned into pGEX (GE Healthcare). These fragments encode the following amino acids of Vkg or Dcg1: VkgN 1–40, VkgC 1505–1761, VkgCN 1520–1612, VkgCCEN 1605–1702, VkgCC 1675–1761, VkgCCEN1 1623–1680, VkgCCEN2 1605–1680, VkgCCEN3 1623–1702, VkgCA 1505–1617 and 1623–1761, DcgN 1–88 and DcgC 1546–1779. The *vkgFL* and *vkgFLA* transgenes were generated by first constructing a full-length cDNA from overlapping RT-PCR products. PCR was used to delete the five amino acids shown in Fig. 1c for the *vkgFLA* transgene. The *vkgFL* and *vkgFLA* cDNAs were amplified by PCR with EcoRI sites on the primers and ligated into the EcoRI site of pCasper4 which has the ubiquitin enhancer and promoter²⁸ cloned as a PCR product into the KpnI site. This ubiquitin enhancer and promoter fragment was PCR amplified with the following primers containing KpnI sites: ubiquitin enhancer 1, GTGGTACCAGATCTTGTCGCGGACGCAGC; and ubiquitin enhancer 2, GTGGTACCTTACTAATTGGTAACAGCGAGTTA.

RNA *in situ* hybridization and immunostaining. For the graphs in Fig. 5e, f, first the proportion of embryos at the onset of gastrulation from *vkg/+* females lacking a transgene with a mutant *Race* expression pattern relative to the total number at this stage was counted. At least 40 embryos at the correct stage were counted for each of three biological repeats of embryo collections and *in situ* hybridizations with the *Race* probe, resulting in an average of 46% with a mutant *Race* expression pattern; the other 54% were indistinguishable from wild type. *Race* expression in embryos at the onset of gastrulation from *vkg/+*; *vkgFL* and *vkg/+*; *vkgFLA* females was scored as narrower, wild type or expanded, and the proportion of each was calculated. The proportion scored as narrower was expressed as a percentage of embryos with mutant *Race* expression typically observed from *vkg/+* females lacking a transgene (that is, 46%), and plotted. Each data bar represents an average from counts of at least 40 embryos at the appropriate stage from each of three independent transgenic lines. The percentage of embryos with expanded *Race* data was calculated in the same way, except that the proportion expanded was expressed relative to the percentage typically observed as wild type in the absence of a transgene (54%).

Immunostaining was performed with the following primary and secondary antibody combinations: mouse monoclonal anti-GFP-20 (Sigma, 1:500) or mouse anti-spectrin (Developmental Studies Hybridoma Bank, 1:20) with anti-mouse-FITC (Jackson, 1:100); rabbit anti-Pat²¹ (1:500) or rabbit anti-Dcg1 (refs 16, 19; 1:1,000) with anti-rabbit-Alexa594 (Invitrogen, 1:500); rabbit anti-Sog4B (ref. 35; 1:500) with anti-rabbit-alkaline-phosphatase (Promega, 1:250); rabbit anti-GFP (Abcam ab290, 1:500) or rabbit anti-pMad³⁶ (1:500) with anti-rabbit-FITC (Jackson, 1:200); and rat anti-Bam³⁷ (1:1,000) with

anti-rat-Alexa594. Dpp-HA was detected in embryos from *vkg/Dpp-HA*; *Dpp-HA/+* females crossed to *Dpp-HA* males using rat monoclonal anti-HA 3F10 (Roche, 1:500) and anti-rat-alkaline-phosphatase (Promega, 1:1,000) as described¹².

***In vitro* pull-down assays.** For the pull-down assays, appropriate amounts of normalized GST-fusion proteins and test protein were incubated in pull-down buffer (200 mM NaCl, 400 mM HEPES, pH 7.9, 50 mM MgCl₂) supplemented with 10 mM EDTA and 10 mM dithiothreitol for 2 h at 4 °C. After extensive washes in pull-down buffer containing 10 mM EDTA, 10 mM dithiothreitol and 0.1% NP40, beads were resuspended in SDS loading buffer, boiled and subjected to western blot analysis. The following antibodies were used: anti-HA (Roche), anti-Flag (Sigma), anti-His (Novagen) and anti-Myc (Santa Cruz). For the release experiment (Fig. 5b), after the washes the appropriate protein was added for 2 h at 4 °C, and then the supernatant was collected after centrifugation. The beads were washed and resuspended in SDS loading buffer.

Surface plasmon resonance. Surface plasmon resonance measurements were performed using a BIAcore 3000 instrument (BIAcore AB). Recombinant Dpp or BMP4 (10 µg ml⁻¹, both R&D systems) diluted in pull-down buffer were immobilized at pH 8 on carboxymethylated dextran surfaces of CM5 sensor chips using amine-coupling chemistry. The CM5 sensor chips were activated with a 1:1 mixture of 0.1 M *N*-hydroxysuccinimide and 0.4 M *N*-ethyl-*N'*-(dimethylaminopropyl)carbodiimide and blotted twice with ethanolamine. Binding assays were performed in pull-down buffer at 25 °C. GST, GST-Vkg and GST-Dcg1 were diluted in pull-down buffer. Human placental collagen IV (BD Biosciences) was diluted in pull-down buffer containing 500 mM NaCl at pH 6. For binding between Dpp and either GST-Vkg or GST-Dcg1, the association was monitored for 6 min at a flow rate of 20 µl min⁻¹, dissociated over 15 min, stabilized for 3 min and then regenerated in 1 M NaCl, 1 mM EDTA. For binding between human collagen IV and BMP4, the association was monitored for 10 min at a flow rate of 20 µl min⁻¹, dissociated over 15 min, stabilized for 3 min and then regenerated in 40 mM NaOH. Each analysis was performed in triplicate. Data analysis was performed using BIAevaluation 4.1 software and the data were fitted to a 1:1 Langmuir binding model with correction for refractive indices differences.

35. Srinivasan, S., Rashka, K. E. & Bier, E. Creation of a Sog morphogen gradient in the *Drosophila* embryo. *Dev. Cell* **2**, 91–101 (2002).
36. Persson, U. et al. The L45 loop in type I receptors for TGF-β family members is a critical determinant in specifying Smad isoform activation. *FEBS Lett.* **434**, 83–87 (1998).
37. McKearin, D. & Ohlstein, B. A role for the *Drosophila* bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. *Development* **121**, 2937–2947 (1995).

LETTERS

Event-horizon-scale structure in the supermassive black hole candidate at the Galactic Centre

Sheperd S. Doeleman¹, Jonathan Weintraub², Alan E. E. Rogers¹, Richard Plambeck³, Robert Freund⁴, Remo P. J. Tilanus^{5,6}, Per Friberg⁵, Lucy M. Ziurys⁴, James M. Moran², Brian Corey¹, Ken H. Young², Daniel L. Smythe¹, Michael Titus¹, Daniel P. Marrone^{7,8}, Roger J. Cappallo¹, Douglas C.-J. Bock⁹, Geoffrey C. Bower³, Richard Chamberlin¹⁰, Gary R. Davis⁵, Thomas P. Krichbaum¹¹, James Lamb¹², Holly Maness³, Arthur E. Niell¹, Alan Roy¹¹, Peter Strittmatter⁴, Daniel Werthimer¹³, Alan R. Whitney¹ & David Woody¹²

The cores of most galaxies are thought to harbour supermassive black holes, which power galactic nuclei by converting the gravitational energy of accreting matter into radiation¹. Sagittarius A* (Sgr A*), the compact source of radio, infrared and X-ray emission at the centre of the Milky Way, is the closest example of this phenomenon, with an estimated black hole mass that is 4,000,000 times that of the Sun^{2,3}. A long-standing astronomical goal is to resolve structures in the innermost accretion flow surrounding Sgr A*, where strong gravitational fields will distort the appearance of radiation emitted near the black hole. Radio observations at wavelengths of 3.5 mm and 7 mm have detected intrinsic structure in Sgr A*, but the spatial resolution of observations at these wavelengths is limited by interstellar scattering^{4–7}. Here we report observations at a wavelength of 1.3 mm that set a size of 37^{+16}_{-10} microarcseconds on the intrinsic diameter of Sgr A*. This is less than the expected apparent size of the event horizon of the presumed black hole, suggesting that the bulk of Sgr A* emission may not be centred on the black hole, but arises in the surrounding accretion flow.

The proximity of Sgr A* makes the characteristic angular size scale of the Schwarzschild radius ($R_{\text{Sch}} = 2GM/c^2$) larger than for any other black hole candidate. At a distance of ~ 8 kpc (ref. 8), the Sgr A* Schwarzschild radius is $10 \mu\text{as}$, or 0.1 astronomical unit (AU). Multi-wavelength monitoring campaigns^{9–11} indicate that activity on scales of a few R_{Sch} in Sgr A* is responsible for observed short-term variability and flaring from radio to X-rays, but direct observations of structure on these scales by any astronomical technique has not been possible. Very-long-baseline interferometry (VLBI) at 7 mm and 3.5 mm wavelength shows the intrinsic size of Sgr A* to have a wavelength dependence, which yields an extrapolated size at 1.3 mm of $20\text{--}40 \mu\text{as}$ (refs 6, 7). VLBI images at wavelengths longer than 1.3 mm, however, are dominated by interstellar scattering effects that broaden images of Sgr A*. Our group has been working to extend VLBI arrays to 1.3 mm wavelength, to reduce the effects of interstellar scattering, and to utilize long baselines to increase angular resolution with a goal of studying the structure of Sgr A* on scales commensurate with the putative event horizon of the black hole. Previous pioneering VLBI work at 1.4 mm wavelength detected Sgr A* on 980-km projected baselines, but calibration

uncertainties resulted in a range for the derived size of $50\text{--}170 \mu\text{as}$ (ref. 12).

On 10 and 11 April 2007, we observed Sgr A* at 1.3 mm wavelength with a three-station VLBI array consisting of the Arizona Radio Observatory 10-m Submillimetre Telescope (ARO/SMT) on Mount Graham in Arizona, one 10-m element of the Combined Array for Research in Millimeter-wave Astronomy (CARMA) in Eastern California, and the 15-m James Clerk Maxwell Telescope (JCMT) near the summit of Mauna Kea in Hawaii. A hydrogen maser time standard and high-speed VLBI recording system were installed at both the ARO/SMT and CARMA sites to support the observation. The JCMT partnered with the Submillimetre Array (SMA) on Mauna Kea, which housed the maser and the VLBI recording system and provided a maser-locked receiver reference to the JCMT. Two 480-MHz passbands sampled to two-bit precision were recorded at each site, an aggregate recording rate of 3.84×10^9 bits per second (Gbit s^{-1}). Standard VLBI practice is to search for detections over a range of interferometer delay and delay rate. Six bright quasars were detected with high signal to noise on all three baselines allowing array geometry, instrumental delays and frequency offsets to be accurately calibrated. This calibration greatly reduced the search space for detections of Sgr A*. All data were processed on the Mark4 correlator at the MIT Haystack Observatory in Massachusetts.

On both 10 and 11 April 2007, Sgr A* was robustly detected on the short ARO/SMT–CARMA baseline and the long ARO/SMT–JCMT baseline. On neither day was Sgr A* detected on the CARMA–JCMT baseline, which is attributable to the sensitivity of the CARMA station being about a third that of the ARO/SMT (owing to weather, receiver temperature and aperture efficiency). Table 1 lists the Sgr A* detections on the ARO/SMT–JCMT baseline. The high signal to noise ratio, coupled with the tight grouping of residual delays and delay rates, makes the detections robust and unambiguous.

There are too few visibility measurements to form an image by the usual Fourier transform techniques; hence, we fit models to the visibilities (shown in Fig. 1). We first modelled Sgr A* as a circular Gaussian brightness distribution, for which one expects a Gaussian relationship between correlated flux density and projected baseline length. The weighted least-squares best-fit model (Fig. 1) corresponds to a Gaussian with total flux density of $2.4 \pm 0.5 \text{ Jy}$ and full

¹Massachusetts Institute of Technology (MIT) Haystack Observatory, Off Route 40, Westford, Massachusetts 01886, USA. ²Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ³University of California Berkeley, Department of Astronomy, 601 Campbell, Berkeley, California 94720-3411 USA. ⁴Arizona Radio Observatory, Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson Arizona 85721-0065, USA. ⁵James Clerk Maxwell Telescope, Joint Astronomy Centre, 660 North A'ohoku Place University Park, Hilo, Hawaii 96720, USA. ⁶Netherlands Organization for Scientific Research, Laan van Nieuw Oost-Indie 300, NL2509 AC The Hague, The Netherlands. ⁷National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22903-2475, USA. ⁸Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637, USA. ⁹CARMA, PO Box 968, Big Pine, California 93513-0968, USA. ¹⁰Caltech Submillimeter Observatory, 111 Nowelo Street, Hilo, Hawaii 96720, USA. ¹¹Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany. ¹²OVRO, California Institute of Technology, 100 Leighton Lane, Big Pine, California 93513-0968, USA. ¹³University of California Berkeley, Space Sciences Laboratory, Berkeley, California 94720-7450, USA.

Table 1 | VLBI detections of Sgr A* on the ARO/SMT–JCMT baseline at 1.3 mm wavelength

Date (UT)	Correlated flux density (Jy)	SNR	Residual delay (ns)	Residual delay rate (ps s ⁻¹)	Projected baseline length (10 ⁶ λ)
10 April 2007 12:20	0.38	5.8	−4.9	−0.29	3,558
11 April 2007 11:00	0.37	5.0	−7.2	−0.25	3,443
11 April 2007 11:40	0.34	5.4	−7.9	−0.21	3,535
11 April 2007 12:00	0.31	5.8	−8.0	−0.19	3,556

Columns are observation date, correlated flux density on the ARO/SMT–JCMT baseline, signal to noise ratio of the VLBI detection, delay and delay-rate residual to the correlator model, and the baseline length projected in the direction of Sgr A*. Each detection was made by incoherently averaging²³ the VLBI signal and searching for a peak in signal to noise ratio over a range of ± 18 ns in delay and ± 2 ps s⁻¹ in delay rate (500 Nyquist sampled search points). For detections on 11 April, data were averaged over 10-min observing scans. The detection on 10 April averaged two 10-min scans together at 12:20 and 12:40 UT to increase integration time. The offset in residual delay between 10 April and 11 April is due to slowly varying instrumental effects and is seen at this same level for nearby quasar calibrators. The statistics of VLBI fringe detection with incoherent averaging are non-Gaussian, and the probability of false detection (the chance a pure noise spike could masquerade as a detection) is a very sharp function of SNR. In the fringe searches on the ARO/SMT–JCMT baseline, for example, SNR of 4.5 is required to give a robust probability of false detection of $< 10^{-6}$, and for SNR of 5.8 in the incoherent fringe search, the probability of false detection is below 10^{-9} . Out of a total of 15 separate 10-min scans, Sgr A* was detected four times on the ARO/SMT–JCMT baseline. Given the strength of these detections, one would expect a higher detection rate than the observed 25%. The low detection rate could be due to intrinsic variations in Sgr A* flux density, but it is more likely to be due to a combination of both pointing errors and variable atmospheric coherence, which would lower fringe search sensitivity, especially at the low elevations at which all sites observed Sgr A*. To convert to Jy, data were calibrated using system temperature, opacity and gain measurements made at all sites.

width at half maximum (FWHM) of 43^{+14}_{-8} μ as where errors are 3σ . On the assumption of a Gaussian profile, the intrinsic size of Sgr A* can be extracted from our measurement assuming that the scatter broadening adds in quadrature with the intrinsic size. At a wavelength of 1.3 mm the scattering size extrapolated from previous longer-wavelength VLBI¹³ is 22 μ as along a position angle 80° degrees east of north on the sky, closely aligned with the orientation of the ARO/SMT–JCMT baseline. Removing the scattering effects results in a 3σ range for the intrinsic size of Sgr A* equal to 37^{+16}_{-10} μ as. The 3σ intrinsic size upper limit at 1.3 mm, combined with a lower limit to the mass of Sgr A* of 4×10^5 solar masses, M_{\odot} , from proper-motion work^{14,15}, yields a lower limit for the mass density of $9.3 \times 10^{22} M_{\odot} \text{pc}^{-3}$. This limit is an order of magnitude larger than previous estimates⁷, and two orders of magnitude below the critical density required for a black hole of $4 \times 10^6 M_{\odot}$. This density lower

limit and central mass would rule out most alternatives to a black hole for Sgr A* because other concentrations of matter would have collapsed or evaporated on timescales that are short compared with the age of the Milky Way¹⁶.

Figure 2 shows both observed and intrinsic sizes for Sgr A* over a wide range of wavelengths along with the scattering model¹³ and the weighted least-squares power-law fit to the intrinsic size measurements. At 1.3 mm wavelength the interstellar scattering size is less than the intrinsic size, demonstrating that VLBI at this wavelength can directly detect structures in Sgr A* on event-horizon scales. The intrinsic size dependence on wavelength, λ^{α} ($\alpha = 1.44 \pm 0.07$, 1σ), confirms that the Sgr A* emission region is stratified, with different wavelengths probing spatially distinct layers. The λ^{α} fit also provides an improved extrapolation to intrinsic sizes at submillimetre wavelengths consistent with emission models that produce X-ray emission

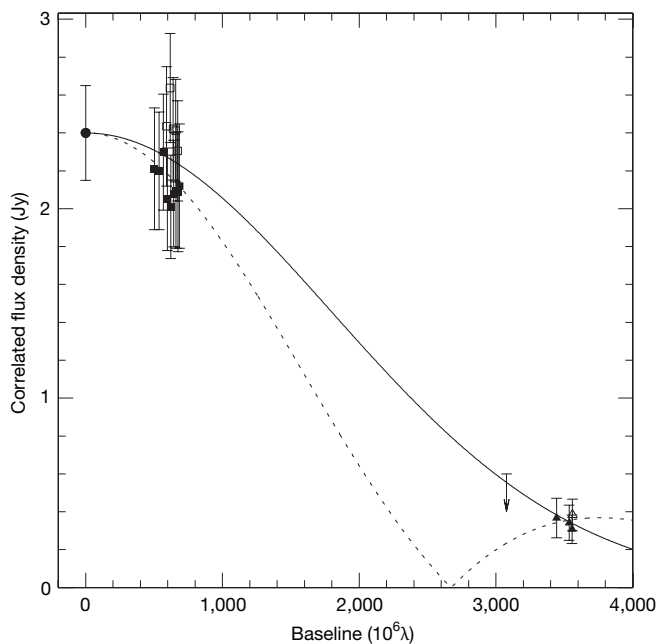


Figure 1 | Fitting the size of Sgr A* with 1.3 mm wavelength VLBI. Shown are the correlated flux density data on the ARO/SMT–CARMA and ARO/SMT–JCMT baselines plotted against projected baseline length (errors are 1σ). Squares show ARO/SMT–CARMA baseline data and triangles show ARO/SMT–JCMT data, with open symbols for 10 April and filled symbols for 11 April. The solid line shows the weighted least-squares best fit to a circular Gaussian brightness distribution, with FWHM size of 43.0 μ as. The dotted line shows a uniform thick-ring model with an inner diameter of 35 μ as and an outer diameter of 80 μ as convolved with scattering effects due to the interstellar medium. The total flux density measurement made with the CARMA array over both days of observing (2.4 ± 0.25 Jy; 1σ) is shown as a filled circle. An upper limit for flux density of 0.6 Jy, derived from non-detections on the JCMT–CARMA baseline, is represented with an arrow near a baseline length of $3,075 \times 10^6 \lambda$.

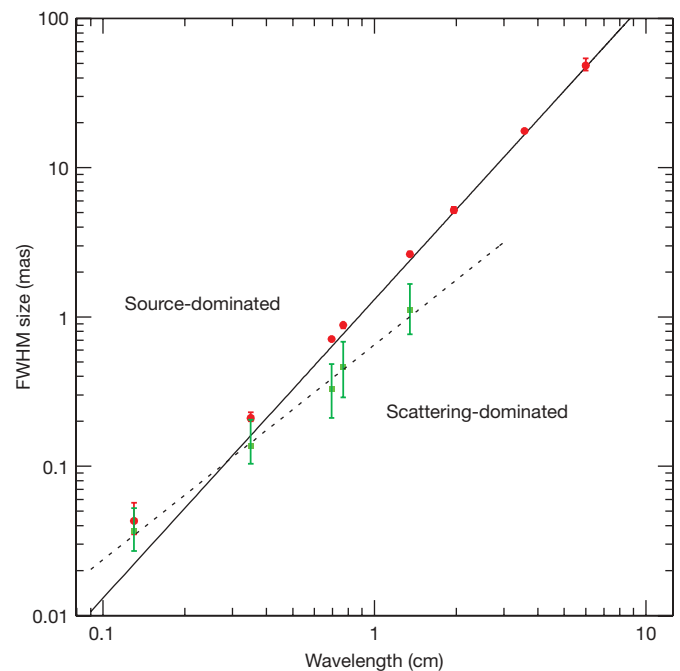


Figure 2 | Observed and intrinsic size of Sgr A* as a function of wavelength. Red circles show major-axis observed sizes of Sgr A* from VLBI observations (all errors 3σ). Data from wavelengths of 6 cm to 7 mm are from ref. 13, data at 3.5 mm are from ref. 7, and data at 1.3 mm are from the observations reported here. The solid line is the best-fit λ^2 scattering law from ref. 13, and is derived from measurements made at $\lambda > 17$ cm. Below this line, measurements of the intrinsic size of Sgr A* are dominated by scattering effects, while measurements that fall above the line indicate intrinsic structures that are larger than the scattering size (a ‘source-dominated’ regime). Green points show derived major-axis intrinsic sizes from $2 \text{ cm} < \lambda < 1.3 \text{ mm}$ and are fitted with a λ^{α} power law ($\alpha = 1.44 \pm 0.07$, 1σ) shown as a dotted line. When the 1.3-mm point is removed from the fit, the power-law exponent becomes $\alpha = 1.56 \pm 0.11$ (1σ).

from inverse Compton scattering of longer-wavelength photons^{9–11}. The minimum intrinsic brightness temperature derived from our 1.3-mm results is 2×10^{10} K.

The data presented here confirm structure in Sgr A* on linear scales of $\sim 4R_{\text{Sch}}$, but the exact nature of this structure is not well determined. The assumption of a Gaussian model above is motivated by simplicity, but the increased angular resolution of VLBI at 1.3 mm will soon allow consideration and testing of more complex structures. As an example, the 1.3-mm VLBI data are also well fitted by a uniform thick ring of inner diameter $35 \mu\text{as}$ and outer diameter $80 \mu\text{as}$ that is convolved with the expected scattering in the interstellar medium (Fig. 1). Such structures are motivated by simulations of the Sgr A* accretion region that use full general relativistic ray tracing^{17,18} and magneto-hydrodynamic effects¹⁹, and which predict a 'shadow' or null in emission in front of the black hole position, especially in the case of face-on accretion disks. The upper limits on correlated flux density from the JCMT–CARMA baseline (Fig. 1) cannot currently discriminate between Gaussian and ring models, but expected and planned increases in both VLBI sensitivity and baseline coverage over the next five years will allow such detailed comparisons.

At present, Sgr A* has been shown to be coincident with the position of the unseen central mass only at the ~ 10 mas level³. It is an open question whether or not the Sgr A* source is centred on the black hole. Indeed, several models predict an offset between Sgr A* and the black hole position. In jet models of Sgr A* (ref. 20), for example, millimetre and submillimetre emission arises at a point in the relativistic plasma stream where the optical depth is close to unity, and the peak in emission can be spatially separated from the black hole. Simulations of accretion disks that are inclined to our line of sight show kinematic (Doppler) brightening on the approaching section of the disk, which also results in an emission peak that is off to one side of the black hole^{17,18,19}. Even for modest accretion disk inclinations, this emission peak can be asymmetric and compact with a morphology dependent on a number of factors including black hole spin, underlying magnetic field structure and inner disk radius.

The intrinsic size derived in this work by fitting the circular Gaussian model can be used to argue that Sgr A* is not a spherically symmetric photosphere centred on the central dark mass. This is because radiation originating from a spherical surface at a given radius from a black hole is strongly lensed by gravity, and presents a larger apparent size to observers on the Earth. Such a surface of radius R centred on a non-rotating black hole will have an apparent radius, R_a , given by^{21,22}

$$R_a = \begin{cases} 3\sqrt{3}R_{\text{Sch}}/2 & \text{if } R \leq 1.5R_{\text{Sch}} \\ R/\sqrt{1-R_{\text{Sch}}/R} & \text{if } R > 1.5R_{\text{Sch}} \end{cases}$$

This has the important consequence that distant observers will measure a minimum apparent diameter of $\sim 5.2 R_{\text{Sch}}$ for all objects centred on the black hole that have radii less than $1.5 R_{\text{Sch}}$ (the minimum circular orbit for photons). In the case of Sgr A*, this corresponds to a minimum apparent diameter of $52 \mu\text{as}$. This size is only marginally consistent with the 3σ upper limit on the intrinsic size derived from our 1.3 mm VLBI observations, and suggests that Sgr A* arises in a region offset from the black hole, presumably in a compact portion of an accretion disk or jet that is Doppler-enhanced by its velocity along our line of sight. This lensing argument also holds in the case of a maximally rotating black hole of the same mass, for which the minimum apparent size in the equatorial plane would be $45 \mu\text{as}$ (ref. 22), which is also larger than the intrinsic size derived here. The intrinsic sizes of Sgr A* measured with VLBI at 3.5 mm and 7 mm exceed the minimum apparent size, and thus cannot similarly be used to constrain the location of Sgr A* relative to the black hole.

Detection of the event-horizon-scale structure reported here indicates that future VLBI observations at $\lambda \leq 1.3$ mm will open a new

window onto fundamental black hole physics through observations of our Galactic Centre. Plans to increase the sensitivity of the VLBI array described here by factors of up to 10 are under way, and the addition of more VLBI stations will increase baseline coverage and the ability to model increasingly complex structures. At projected VLBI array sensitivities, Sgr A* will be detected on multiple baselines within 10-s timescales, enabling sensitive tests for time-variable structures such as those suggested by orbiting hotspot¹⁸ and flaring models^{9–11}.

Received 6 May; accepted 8 July 2008.

1. Kormendy, J. & Richstone, D. Inward bound: the search for supermassive black holes in galactic nuclei. *Annu. Rev. Astron. Astrophys.* **33**, 581–624 (1995).
2. Schödel, R. et al. A star in a 15.2-year orbit around the supermassive black hole at the centre of the Milky Way. *Nature* **419**, 694–696 (2002).
3. Ghez, A. et al. Stellar orbits around the Galactic Center black hole. *Astrophys. J.* **620**, 744–757 (2005).
4. Rogers, A. E. E. et al. Small-scale structure and position of Sagittarius A* from VLBI at 3 millimeter wavelength. *Astrophys. J.* **434**, L59–L62 (1994).
5. Doeleman, S. S. et al. Structure of Sagittarius A* at 86 GHz using VLBI closure quantities. *Astron. J.* **121**, 2610–2617 (2001).
6. Bower, G. C. et al. Detection of the intrinsic size of Sagittarius A* through closure amplitude imaging. *Science* **304**, 704–708 (2004).
7. Shen, Z. Q., Lo, K. Y., Liang, M.-C., Ho, P. T. P. & Zhao, J.-H. A size of ~ 1 AU for the radio source Sgr A* at the centre of the Milky Way. *Nature* **438**, 62–64 (2005).
8. Reid, M. J. The distance to the center of the Galaxy. *Annu. Rev. Astron. Astrophys.* **31**, 345–372 (1993).
9. Marrone, D. et al. An X-ray, IR, and submillimeter flare of Sagittarius A*. *Astrophys. J.* **682**, 373–383 (2008).
10. Yusef-Zadeh, F. et al. A multiwavelength study of Sgr A*: The role of near-IR flares in production of X-ray, soft γ -ray, and submillimeter emission. *Astrophys. J.* **644**, 198–213 (2006).
11. Eckart, A. et al. The flare activity of Sagittarius A*. New coordinated mm to X-ray observations. *Astron. Astrophys.* **450**, 535–555 (2006).
12. Krichbaum, T. P. et al. VLBI observations of the Galactic Center source Sgr A* at 86 GHz and 215 GHz. *Astron. Astrophys.* **335**, L106–L110 (1998).
13. Bower, G. C., Goss, W. M., Falcke, H., Backer, D. C. & Lithwick, Y. The intrinsic size of Sagittarius A* from 0.35 to 6 cm. *Astrophys. J.* **648**, L127–L130 (2006).
14. Backer, D. C. & Sramek, R. A. Proper motion of the compact, nonthermal radio source in the Galactic Center, Sagittarius A*. *Astrophys. J.* **524**, 805–815 (1999).
15. Reid, M. J. & Brunthaler, A. The proper motion of Sagittarius A*. II. The mass of Sagittarius A*. *Astrophys. J.* **616**, 872–884 (2004).
16. Maoz, E. Dynamical constraints on alternatives to supermassive black holes in galactic nuclei. *Astrophys. J.* **494**, L181–L184 (1998).
17. Falcke, H., Melia, F. & Agol, E. Viewing the shadow of the black hole at the Galactic Center. *Astrophys. J.* **528**, L13–L16 (2000).
18. Broderick, A. E. & Loeb, A. Imaging optically-thin hotspots near the black hole horizon of Sgr A* at radio and near-infrared wavelengths. *Mon. Not. R. Astron. Soc.* **367**, 905–916 (2006).
19. Noble, S. C., Leung, P. K., Gammie, C. F. & Book, L. G. Simulating the emission and outflows from accretion discs. *Class. Quant. Gravity* **24**, S259–S274 (2007).
20. Falcke, H. & Markoff, S. The jet model for Sgr A*: Radio and X-ray spectrum. *Astron. Astrophys.* **362**, 113–118 (2000).
21. Broderick, A. E. & Narayan, R. On the nature of the compact dark mass at the galactic center. *Astrophys. J.* **638**, L21–L24 (2006).
22. Frolov, V. P. & Novikov, I. D. *Black Hole Physics: Basic Concepts and New Developments* (Kluwer, 1998).
23. Rogers, A. E. E., Doeleman, S. S. & Moran, J. M. Fringe detection methods for very long baseline arrays. *Astron. J.* **109**, 1391–1401 (1995).

Acknowledgements High-frequency VLBI work at MIT Haystack Observatory is supported by grants from the National Science Foundation. The Submillimeter Array is a joint project between the Smithsonian Astrophysical Observatory and the Academia Sinica Institute of Astronomy and Astrophysics. We thank G. Weaver for the loan of a frequency reference from Johns Hopkins University Applied Physics Labs; J. Davis for use of GPS equipment; I. Diegel, R. Vessot, D. Phillips and E. Mattison for assistance with hydrogen masers; the NASA Geodesy Program for loan of the CARMA Hydrogen Maser; D. Kubo, J. Test, P. Yamaguchi, G. Reiland, J. Hoge and M. Hodges for technical assistance; M. Gurwell for SMA calibration data; A. Kerr and A. Lichtenberger for contributions at ARO/SMT; A. Broderick, V. Fish, A. Loeb and I. Shapiro for discussions; and the staff at all participating facilities.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.S.D. (sdoeleman@haystack.mit.edu).

Electronic structure of the iron-based superconductor LaOFeP

D. H. Lu¹, M. Yi¹, S.-K. Mo^{1,2}, A. S. Erickson³, J. Analytis³, J.-H. Chu³, D. J. Singh⁴, Z. Hussain², T. H. Geballe³, I. R. Fisher³ & Z.-X. Shen¹

The recent discovery of superconductivity in the iron oxypnictide family of compounds^{1–9} has generated intense interest. The layered crystal structure with transition-metal ions in planar square-lattice form and the discovery of spin-density-wave order near 130 K (refs 10, 11) seem to hint at a strong similarity with the copper oxide superconductors. An important current issue is the nature of the ground state of the parent compounds. Two distinct classes of theories, distinguished by the underlying band structure, have been put forward: a local-moment antiferromagnetic ground state in the strong-coupling approach^{12–17}, and an itinerant ground state in the weak-coupling approach^{18–22}. The first approach stresses on-site correlations, proximity to a Mott-insulating state and, thus, the resemblance to the high-transition-temperature copper oxides, whereas the second approach emphasizes the itinerant-electron physics and the interplay between the competing ferromagnetic and antiferromagnetic fluctuations. The debate over the two approaches is partly due to the lack of conclusive experimental information on the electronic structures. Here we report angle-resolved photoemission spectroscopy (ARPES) of LaOFeP (superconducting transition temperature, $T_c = 5.9$ K), the first-reported iron-based superconductor². Our results favour the itinerant ground state, albeit with band renormalization. In addition, our data reveal important differences between these and copper-based superconductors.

In Fig. 1 we compare the angle-integrated photoemission spectrum (AIPES) with the density of states obtained from the local-density-approximation (LDA) band structure calculations. It is important to note that the peak near the Fermi level (E_F) is as strong as the valence band peak, in sharp contrast with the typical valence band spectrum of copper oxide superconductors, as shown in the inset of Fig. 1a. The valence band spectrum of copper oxide superconductors is characterized by a weak feature near E_F on top of a broad valence band peak, consistent with the doped-Mott-insulator picture. This clear disparity between the iron-based superconductor and the copper oxide superconductors suggests that itinerant-electron physics rather than Mott physics is a more appropriate starting point for the iron-based superconductors, at least for LaOFeP. Our data also disagree with some recent AIPES data^{23,24} obtained from polycrystalline samples that show only a very small peak near E_F on top of a large valence band peak, which is reminiscent of the valence band spectra of copper oxide superconductors. This difference may be due to the surface quality of polycrystalline samples, as is often the case for oxides²⁵. On balance, our data do not support theoretical models assuming strongly antiferromagnetic ground states (at least not those currently being formulated, albeit for the LaOFeAs system^{12–15}), as there is no evidence in our spectra for exchange splitting

of the iron d -electron states, and agreement between our valence band spectrum and the density of states calculated using such models is poorer in comparison with the density of states calculated in the LDA assuming an itinerant ground state.

More detailed information can be obtained from angle-resolved data. To understand the seemingly complex multiband electronic structure, we superimpose the LDA band structures on top of our data (Fig. 2). A quantitative agreement can be found between the angle-resolved photoemission spectra and the calculated band dispersions after shifting the calculated bands up by 0.11 eV and then renormalizing by a factor of 2.2. Note that the values of the E_F shift and the band renormalization factor are chosen to obtain the best match of the two higher binding energy bands at the Γ point. Although the renormalized bands using this set of parameters fit the bands near Γ very well, the match near the X point and the M point is less perfect. This suggests that different bands may have slightly different renormalization effects. Nevertheless, the overall level of agreement between the experiments and the calculations is significant, as nearly all features in our data have corresponding bands in the calculations, indicating that the LDA with the assumption of an itinerant ground state captures the essence of the electronic structure of this system. This again suggests that the iron-based superconductors, or at least LaOFeP, are different from copper oxide superconductors. We also note that the measured dispersions show no similarity with the band structure calculations of LaOFeAs calculated assuming an antiferromagnetic ground state^{12–15}.

To extract more information from angle-resolved photoemission spectra, a simple analysis of momentum distribution curves is done for the high-symmetry cuts. A Fermi velocity (v_F) of 1.0 ± 0.2 eV \AA^{-1} (equivalent to $(1.5 \pm 0.3) \times 10^5$ m s⁻¹) is obtained for all three bands individually. For comparison, the values extracted from the LDA calculations, after taking into account the E_F shift, are 1.5 or 1.7, 1.4, and 2.4 or 3.5 eV \AA^{-1} for the Γ_1 , Γ_2 and M bands, respectively. Note that two different numbers are given for both the Γ_1 band and the M band because each contains two nearly degenerate bands. This observation demonstrates that the renormalization effects are different for different bands, as anticipated above, indicating that correlation effects are appreciable and not isotropic. However, these v_F renormalization values as well as the total-bandwidth rescaling factor of 2.2 are comparable to those of Sr₂RuO₄, which is a correlated Fermi liquid and is reasonably well described by theories using itinerant band structure as the starting point²⁶. The corresponding electron-band masses m^* extracted from our data are, in units of the free electron mass, 1.4 ± 0.3 , 4.6 ± 0.5 and 1.3 ± 0.3 for the Γ_1 , Γ_2 and M bands, respectively. We note that the magnetic susceptibility enhancement compared with the bare-band-structure density of

¹Department of Physics, Department of Applied Physics and Stanford Synchrotron Radiation Laboratory, Stanford University, Stanford, California 94305, USA. ²Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ³Geballe Laboratory for Advanced Materials and Department of Applied Physics, Stanford University, Stanford, California 94305-4045, USA. ⁴Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6114, USA.

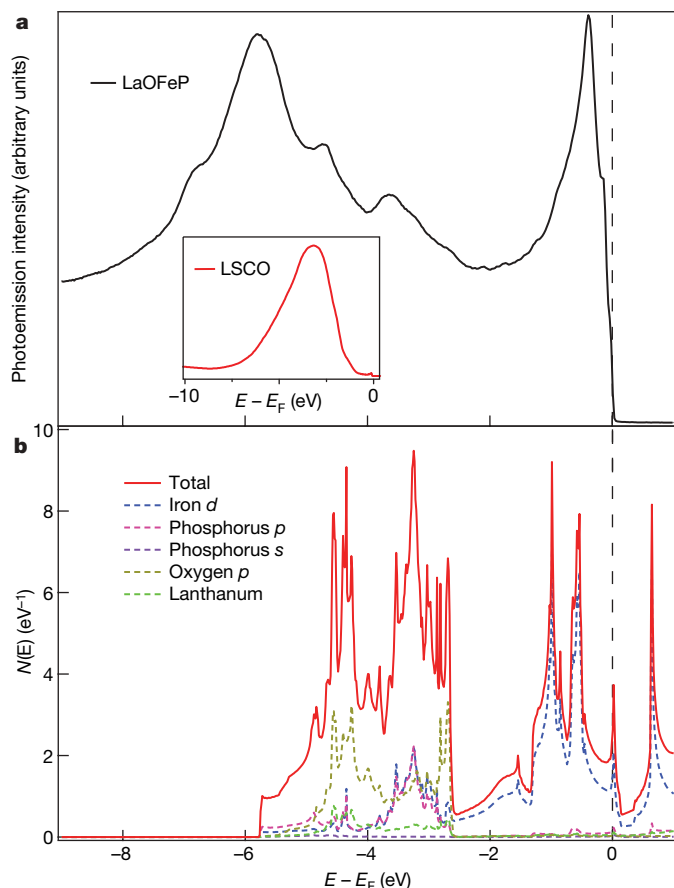


Figure 1 | Comparison between angle-integrated photoemission spectrum and calculated density of states. **a**, Valence band spectrum of LaOFeP taken with 42.5-eV photons using transmission mode (see Supplementary Information). It consists of a sharp, intense peak near the Fermi level that is separated from a number of broad peaks at higher binding energy. The inset shows the valence band of $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (LSCO), for comparison. **b**, LDA density of states and projections onto the linear-augmented-plane-wave spheres. According to the LDA calculations, the near- E_F states have dominant iron d -state character, whereas the peaks at higher binding energy are mixtures of oxygen p states and hybridized iron d and phosphorus p states. In comparison with the calculated density of states, the near- E_F peak has a narrower width than the calculated iron d states and is pushed closer to E_F , which is consistent with the band renormalization effect discussed in Fig. 2. The valence band peaks at higher binding energy, however, are shifted towards higher binding energy, resulting in slightly larger total valence band width.

states is a factor of almost six¹⁸, indicating either a lower-energy-scale renormalization or a strong Stoner renormalization. In this regard, we do not observe any apparent low-energy kink in the dispersion near 50 meV, which is a universal feature in copper oxide superconductors²⁷.

In Fig. 3 we display the energy distribution curves (EDCs) along the same high-symmetry cuts as shown in Fig. 2. Close inspection of these shows that there is no evident pseudogap within our experimental uncertainty in all three bands crossing the Fermi level, in contrast to the ubiquitous pseudogap observed in underdoped copper oxides. The absence of the pseudogap, therefore, marks an important difference between this new iron-based superconductor and copper oxide superconductors. This finding contradicts the recent report of a 20-meV pseudogap in LaOFeP from AIPES²⁸. The difference can be attributed either to the polycrystalline samples used for that measurement having poor surface quality (previous work indicates potential problems associated with impurities²⁵), or to distortion of the AIPES result by states away from the Fermi crossing (k_F). Angle-resolved photoemission spectroscopy of single-crystalline samples is much

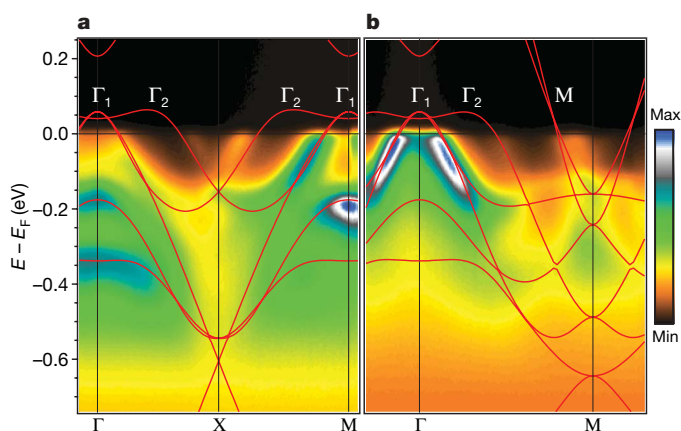


Figure 2 | Comparison between angle-resolved photoemission spectra and LDA band structures along two high-symmetry lines. ARPES data from LaOFeP (image plots) were recorded using 42.5-eV photons with an energy resolution of 16 meV and an angular resolution of 0.3° . For better comparison with experimental data, the LDA band structures using the experimental lattice parameters with relaxed internal atomic positions (see Supplementary Information) are shifted up by ~ 0.11 eV and then renormalized by a factor of 2.2 (red lines). **a**, Along the Γ -X direction, two bands crossing E_F can be clearly identified: one near the Γ point (Γ_1) and one near the X point (Γ_2). These two crossings are associated with two hole-like Fermi surface pockets centred at Γ . According to the LDA calculations, the inner pocket originates from iron d_{xz} and d_{yz} bands that are degenerate at Γ , and the splitting of these two bands close to Γ is too small to be resolved in our data. However, we do see evidence for the splitting at higher binding energy. The outer pocket is derived from the iron $d_{3z^2-r^2}$ states that hybridize with the phosphorus p orbitals and lanthanum orbitals. **b**, Along the Γ -M direction, three E_F crossings are observed in total. In addition to the two crossings associated with two hole pockets, a crossing near the M point can be observed, although the corresponding crossing in the second zone is too weak to be seen, owing to the matrix element effect. This crossing is related to the electron pocket centred at M. The LDA calculations also predict two bands crossing E_F around the M point, which cannot be clearly resolved in our data.

better suited to addressing the pseudogap issue by directly measuring the states near k_F . The same AIPES experiment²⁸ also indicated pseudogap effects with energy scales of 20 and 100 meV in polycrystalline $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ compounds, whereas another AIPES experiment²⁴ found a pseudogap of 15–20 meV in the same polycrystalline compounds. We cannot rule out the possibility of a pseudogap in arsenic-based compounds, which exhibit a spin-density-wave order in their parent compound LaOFeAs, as indicated in neutron scattering studies^{10,11}. However, the similarly observed 20-meV pseudogap in polycrystalline samples of both LaOFeP and LaOFeAs (ref. 28) leads us to suggest a careful re-examination as soon as single crystals of the arsenic-based compounds become available.

Finally, we consider the Fermi surface topology (Fig. 4). Three sheets of Fermi surfaces are clearly observed: two hole pockets centred at Γ and one electron pocket centred at M. Keeping in mind the nearly degenerate Γ_1 and M bands, the observed Fermi surface topology is consistent with the five sheets of Fermi surfaces predicted in band structure calculations²⁹. We note that the outer hole pocket Γ_2 originates from the hybridized iron $d_{3z^2-r^2}$ and phosphorus p states, which have strong k_z dispersion. The topology of this Fermi surface sheet is sensitive to the position of the phosphorus atoms, that is, the level of hybridization, and changes significantly upon doping. Calculating the Fermi surface volume enclosed by the three pockets yields respective electron counts of 1.94, 1.03 and 0.05 for the Γ_1 , Γ_2 and M pockets. Taking into account the unresolved, nearly degenerate sheets under the Γ_1 and M pockets, a total electron count of 5.0 ± 0.1 is obtained, which is smaller than the expected value of 6. This is consistent with the need to shift E_F in order to produce the best fits of the dispersion in Fig. 2. It is too early to be certain how much of

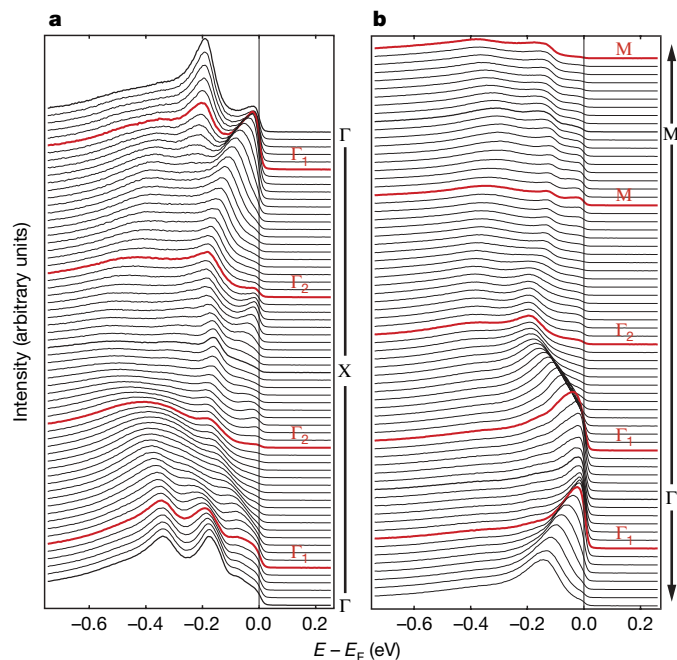


Figure 3 | Energy distribution curves along two high-symmetry lines. **a**, EDCs along the Γ -X direction. **b**, EDCs along the Γ -M direction. EDCs at k_F are plotted in red. The leading-edge midpoints of the red EDCs apparently reach E_F for all bands crossing E_F , indicating, within our experimental uncertainty, the absence of a pseudogap in this system.

this discrepancy is caused by a change in the surface doping, k_z dispersion or subtle surface-structure distortion, which can be significant for the Γ_2 band.

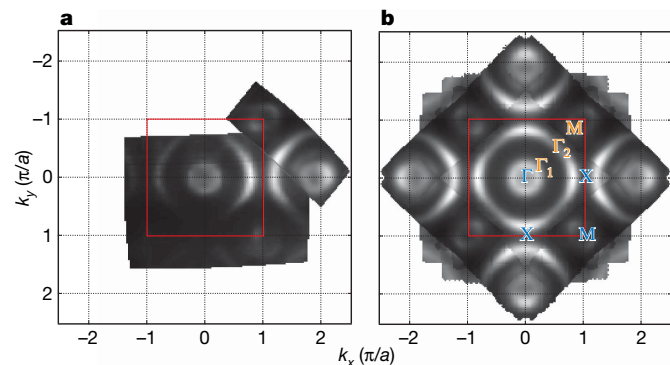


Figure 4 | Fermi surface maps of LaOFeP. **a**, Two sets of Fermi surface mapping (unsymmetrized raw data) are overlaid: the first set covers more than one Brillouin zone and the second set, taken mostly in the second Brillouin zone, yields a better view of the Fermi surface pocket at the M point, which is not well resolved in the first set owing to the polarization issue. The map is obtained by integrating the EDCs over an energy window of $E_F \pm 15$ meV. The red square highlights the boundary of the first Brillouin zone, where a is the in-plane lattice constant. **b**, Symmetrized Fermi surface map obtained by flipping and rotating the raw data shown in **a** along the high-symmetry lines to reflect the symmetry of the crystal structure. We use the Brillouin zone corresponding to the two-iron unit cell with the M point at (π, π) , which is $(\pi, 0)$ in the large Brillouin zone for a simple iron square lattice. Three sheets of Fermi surfaces, labelled Γ_1 , Γ_2 and M, are clearly observed. As discussed above (Fig. 2), Γ_1 , the inner hole pocket observed in our data, should contain two nearly degenerate sheets, and the same is true for the electron pocket around M. Therefore, our data are consistent with the five sheets of Fermi surfaces predicted in band structure calculations²⁹: two hole pockets around Γ , two electron pockets around M, and one highly three-dimensional hole pocket centred at Z.

Careful examination of the data in Fig. 3 reveals another possible discrepancy in band structure comparison, namely a very weak feature around -0.07 eV near Γ (Fig. 3a) that does not seem to have a corresponding band in LDA calculations. Further investigations are required to clarify its origin. Despite these disagreements, all the expected Fermi surface pieces are observed and are in good agreement with experiments in terms of the Brillouin zone locations and signs (hole versus electron). Furthermore, the measured main dispersions agree with the calculated band structures in great detail, as shown in Fig. 2. These observations make a strong case that the itinerant band structure captures the essence of the electronic structure of LaOFeP.

In summary, our ARPES data from LaOFeP suggest that the electronic structure of this material can be described using an itinerant band approach. In comparison with copper oxide superconductors, it has three important contrasting features: it has a much higher density of states near the Fermi level; it has multiple bands and Fermi surface sheets; and it shows no apparent evidence of the pseudogap effect.

Received 31 May; accepted 11 July 2008.

- Day, C. New family of quaternary iron-based compounds superconducts at tens of kelvin. *Phys. Today* **61**, 11–12 (2008).
- Kamihara, Y. *et al.* Iron-based layered superconductor: LaOFeP. *J. Am. Chem. Soc.* **128**, 10012–10013 (2006).
- Kamihara, Y., Watanabe, T., Hirano, M. & Hosono, H. Iron-based layered superconductor $\text{La}[\text{O}_{1-x}\text{F}_x]\text{FeAs}$ ($x = 0.05$ – 0.12) with $T_c = 26$ K. *J. Am. Chem. Soc.* **130**, 3296–3297 (2008).
- Takahashi, H. *et al.* Superconductivity at 43 K in an iron-based layered compound $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$. *Nature* **453**, 376–378 (2008).
- Chen, X. H. *et al.* Superconductivity at 43 K in $\text{SmFeAsO}_{1-x}\text{F}_x$. *Nature* **453**, 761–762 (2008).
- Chen, G. F. *et al.* Superconductivity at 41 K and its competition with spin-density-wave instability in layered $\text{CeO}_{1-x}\text{F}_x\text{FeAs}$. *Phys. Rev. Lett.* **100**, 247002 (2008).
- Ren, Z.-A. *et al.* Superconductivity at 52 K in iron-based F-doped layered quaternary compound $\text{Pr}[\text{O}_{1-x}\text{F}_x]\text{FeAs}$. Preprint at (<http://arxiv.org/abs/0803.4283>) (2008).
- Ren, Z.-A. *et al.* Superconductivity at 55 K in iron-based F-doped layered quaternary compound $\text{Sm}[\text{O}_{1-x}\text{F}_x]\text{FeAs}$. *Chin. Phys. Lett.* **25**, 2215–2216 (2008).
- Wen, H.-H., Mu, G., Fang, L., Yang, H. & Zhu, X. Superconductivity at 25 K in hole doped $(\text{La}_{1-x}\text{Sr}_x)\text{OFeAs}$. *Europhys. Lett.* **82**, 17009 (2008).
- de la Cruz, C. *et al.* Magnetic order close to superconductivity in the iron-based layered $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$ systems. *Nature* **453**, 899–902 (2008).
- Qiu, Y. *et al.* Neutron scattering study of the oxypnictide superconductor $\text{La}(\text{O},\text{F})\text{FeAs}$. Preprint at (<http://arxiv.org/abs/0805.1062>) (2008).
- Cao, C., Hirschfeld, P. J. & Cheng, H. P. Proximity of antiferromagnetism and superconductivity in $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$: effective Hamiltonian from ab initio studies. *Phys. Rev. B* **77**, 220506 (2008).
- Haule, K., Shim, J. H. & Kotliar, G. Correlated electronic structure of $\text{LaO}_{1-x}\text{F}_x\text{FeAs}$. *Phys. Rev. Lett.* **100**, 226402 (2008).
- Yin, Z. P. *et al.* Electron-hole symmetry and magnetic coupling in antiferromagnetic LaOFeAs. *Phys. Rev. Lett.* **101**, 047001 (2008).
- Ma, F., Lu, Z.-Y. & Xiang, T. Antiferromagnetic superexchange interactions in LaOFeAs. Preprint at (<http://arxiv.org/abs/0804.3370>) (2008).
- Fang, C., Yao, H., Tsai, W.-F., Hu, J. & Kivelson, S. Theory of electron nematic order in LaFeAsO. *Phys. Rev. B* **77**, 224509 (2008).
- Xu, C., Mueller, M. & Sachdev, S. Ising and spin orders in iron-based superconductors. *Phys. Rev. B* **78**, 020501 (2008).
- Singh, D. J. & Du, M.-H. Density functional study of $\text{LaFeAsO}_{1-x}\text{F}_x$: a low carrier density superconductor near itinerant magnetism. *Phys. Rev. Lett.* **100**, 237003 (2008).
- Mazin, I. I., Singh, D. J., Johannes, M. D. & Du, M. H. Unconventional sign-reversing superconductivity in $\text{LaFeAsO}_{1-x}\text{F}_x$. *Phys. Rev. Lett.* (in the press); preprint at (<http://arxiv.org/abs/0803.2740>) (2008).
- Zhang, H.-J., Xu, G., Dai, X. & Fang, Z. Enhanced orbital degeneracy in momentum space for LaOFeAs. Preprint at (<http://arxiv.org/abs/0803.4487>) (2008).
- Raghu, S., Qi, X.-L., Liu, C.-X., Scalapino, D. & Zhang, S.-C. Minimal two-band model of the superconducting iron oxypnictides. *Phys. Rev. B* **77**, 220503 (2008).
- Lee, P. A. & Wen, X.-G. Spin-triplet p-wave pairing in a 3-orbital model for FeAs superconductors. Preprint at (<http://arxiv.org/abs/0804.1739>) (2008).
- Ou, H. W. *et al.* Angle integrated photoemission study of $\text{SmO}_{0.85}\text{F}_{0.15}\text{FeAs}$. *Chin. Phys. Lett.* **25**, 2225–2227 (2008).
- Sato, T. *et al.* Superconducting gap and pseudogap in iron-based layered superconductor $\text{La}(\text{O}_{1-x}\text{F}_x)\text{FeAs}$. *J. Phys. Soc. Jpn* **77**, 063708 (2008).
- Arko, A. *et al.* Large, dispersive photoelectron Fermi edge and the electronic structure of $\text{YBa}_2\text{Cu}_3\text{O}_{6.9}$ single crystals measured at 20 K. *Phys. Rev. B* **40**, 2268–2277 (1989).

26. Mackenzie, A. & Maeno, Y. The superconductivity of Sr_2RuO_4 and the physics of spin-triplet pairing. *Rev. Mod. Phys.* **75**, 657–712 (2003).
27. Lanzara, A. *et al.* Evidence for ubiquitous strong electron–phonon coupling in high-temperature superconductors. *Nature* **412**, 510–514 (2001).
28. Ishida, Y. *et al.* Evidence for pseudogap evolutions in high- T_c iron oxypnictides. Preprint at (<http://arxiv.org/abs/0805.2647>) (2008).
29. Lebegue, S. Electronic structure and properties of the Fermi surface of the superconductor LaOFeP . *Phys. Rev. B* **75**, 035110 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Cox, S. M. Kauzlarich and H. Hope for single-crystal X-ray diffraction measurements, and H. Yao, S. A. Kivelson, R. M. Martin, S. C. Zhang and X. L. Qi for discussions. ARPES experiments were performed at the Advanced Light Source, which is operated by the US Department of Energy Office of Basic Energy Science. Work at Stanford and Oak Ridge National Laboratory was supported by the Office of Basic Energy Science, Division of Materials Science and Engineering.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to Z.-X.S. (zxshen@stanford.edu) or D.H.L. (dhlu@slac.stanford.edu).

Nanoscale double emulsions stabilized by single-component block copolypeptides

Jarrold A. Hanson¹, Connie B. Chang², Sara M. Graves², Zhibo Li¹, Thomas G. Mason^{2,3,4} & Timothy J. Deming^{1,2,4}

Water-in-oil-in-water emulsions are examples of double emulsions, in which dispersions of small water droplets within larger oil droplets are themselves dispersed in a continuous aqueous phase^{1–3}. Emulsions occur in many forms of processing and are used extensively by the foods, cosmetics and coatings industries. Because of their compartmentalized internal structure, double emulsions can provide advantages over simple oil-in-water emulsions for encapsulation, such as the ability to carry both polar and non-polar cargos, and improved control over release of therapeutic molecules^{4–6}. The preparation of double emulsions typically requires mixtures of surfactants for stability; the formation of double nanoemulsions, where both inner and outer droplets are under 100 nm, has not yet been achieved^{7–9}. Here we show that water-in-oil-in-water double emulsions can be prepared in a simple process and stabilized over many months using single-component, synthetic amphiphilic diblock copolypeptide surfactants. These surfactants even stabilize droplets subjected to extreme flow, leading to direct, mass production of robust double nanoemulsions that are amenable to nanostructured encapsulation applications in foods, cosmetics and drug delivery.

Although they offer certain advantages over ordinary oil-in-water emulsions, stable water-in-oil-in-water (WOW) emulsions generally do not form spontaneously with a single surfactant and standard emulsification methods^{7,10}. Microfluidics can be used to make double emulsions that are micrometres in size and highly uniform^{8,9}, yet the throughput can be low compared with commercial processes for making polydisperse single emulsions¹¹. Typical methods for making WOW emulsions use a two-step process of first forming an ‘inverse’ water-in-oil emulsion, followed by emulsification of this mixture in water using a combination of surfactants^{2,3,7,12,13}. This process allows control of both droplet volumes if the emulsions are made monodisperse³, yet cannot form stable nanoscale droplets and requires a difficult search for surfactant combinations that can coexist without destabilizing either inner or outer droplet interfaces². Consequently, improving stability and reducing droplet sizes are the key challenges in the development of double emulsions for applications¹⁴.

The block copolypeptide surfactants we designed have the general structure poly(L-lysine·HBr)_x-b-poly(racemic-leucine)_y, K_x(rac-L)_y, where *x* ranged from 20 to 100, and *y* ranged from 5 to 30 residues (Fig. 1a, Supplementary Information). The hydrophilic poly(L-lysine·HBr) segments are highly charged at neutral pH, provide good water solubility¹⁵ and possess abundant amine groups for chemical functionalization¹⁶. Unlike hydrophobic segments of other polymeric amphiphiles, poly(L-leucine) segments adopt rod-like α -helical conformations that give rise to strong interchain associations and poor solubility in common organic solvents¹⁷. Block copolymers of the structure K_xL_y (for example K₆₀L₂₀, Fig. 1b) associate strongly in water to form membranes through packing of the hydrophobic

segments¹⁸. Consequently, we focused on poly(*rac*-leucine) because its disordered chain conformation improves solubility (Supplementary Table 1)^{19,20} and helps to promote surface activity (Supplementary Table 1), and its peptidic nature allows for additional mechanical stabilization of droplet interfaces through interchain hydrogen-bonding in the oil phase²¹.

We screened diblock copolypeptides for emulsification activity by adding silicone oil to aqueous K_x(rac-L)_y solutions (Supplementary Table 1, Supplementary Figs 2a–c, 5a). The resulting mixtures were sheared using a hand-held rotary homogenizer and then passed six times through a high-pressure microfluidic homogenizer (Fig. 1c). All K_x(rac-L)_y samples gave stable WOW nanoemulsions that did not ripen (that is, coarsen in size) or phase-separate for over nine months. Only copolypeptides with low hydrophobic content, for example K₄₀(rac-L)₅, gave emulsions that slowly phase-separated after one year. Other methods of mixing, including ultrasonic

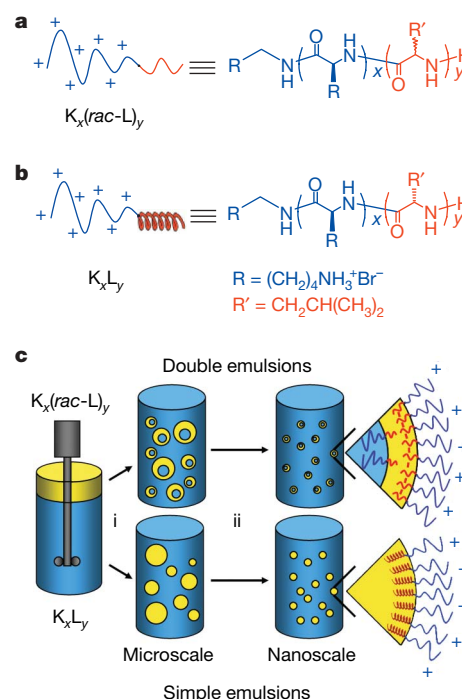


Figure 1 | Structures of block copolypeptide surfactants and emulsification procedure. a, K_x(rac-L)_y. b, K_xL_y. c, Emulsification procedure used to generate both simple and double emulsions. Step (i), ultrasonic or hand-held homogenization; step (ii), microfluidic homogenization. Yellow represents the oil phase, blue the aqueous phase containing block copolypeptide surfactant.

¹Bioengineering Department, ²Department of Chemistry and Biochemistry, ³Department of Physics and Astronomy, and ⁴California NanoSystems Institute, University of California, Los Angeles, California 90095, USA.

mixing, also provided stable emulsions, but with droplets up to several micrometres in diameter (Fig. 1c). Use of hydrophobic segments longer than 30 residues greatly diminished aqueous solubility (Supplementary Table 1); for instance, $K_{40}(\text{rac-L})_{30}$ could only be dissolved up to 1 mM. As controls, we also used 0.1 mM suspensions of $K_{60}L_{20}$ and K_{60} as surfactants: $K_{60}L_{20}$ did form stable emulsions and K_{60} failed to emulsify oil and water mixtures (Supplementary Fig. 4). These results indicated that $K_x(\text{rac-L})_y$ surfactants give stable emulsions over a broad range of compositions and concentrations.

To probe droplet structure, we imaged block-copolypeptide-stabilized emulsions by using optical microscopy and cryogenic transmission electron microscopy (CTEM). All samples with $K_x(\text{rac-L})_y$ were found to contain oil droplets, each containing predominately a single internal aqueous droplet with consistent inner to outer volume ratios (Fig. 2a, Supplementary Figs 2, 3). In contrast, the emulsions formed using $K_{60}L_{20}$ contained only simple oil droplets (Fig. 2b), revealing that the racemic-leucine segments play a key part in stabilizing the double emulsion structure. As copolypeptide hydrophobic content was decreased, droplet sizes increased (Supplementary Table 1, Supplementary Fig. 5c), suggesting that copolymer composition influences interfacial mean curvature. Average droplet diameters also increased when the concentration of $K_{40}(\text{rac-L})_{20}$ was decreased (Supplementary Fig. 5a). Likewise, decreasing the volume fraction of oil yielded smaller emulsion droplets (Supplementary Fig. 5b). Emulsions always formed such that water remained the continuous phase and did not invert up to oil volume fractions approaching 50%. In addition to polydimethylsiloxane (PDMS), other immiscible liquids such as dodecane, soybean oil and methyl oleate gave emulsions using 1 mM $K_{40}(\text{rac-L})_{20}$ in water. The versatility of our design was shown by formation of stable emulsions using $R_{40}(\text{rac-L})_{10}$ or $E_{40}(\text{rac-L})_{10}$, containing guanidinium or carboxylate functionality of L-arginine (R) and L-glutamate (E), respectively (Supplementary Fig. 3a, b).

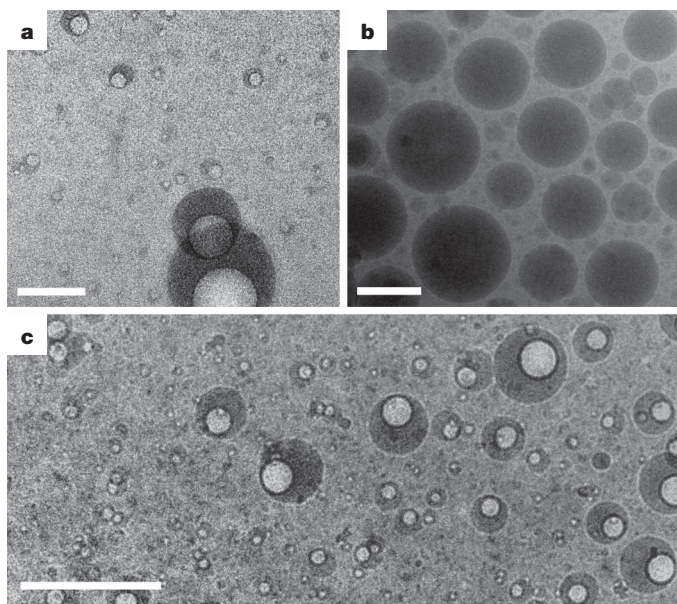


Figure 2 | Cryogenic transmission electron microscopy of copolypeptide-stabilized emulsions prepared using a microfluidic homogenizer. Vitriified water gives a light background and silicone oil appears dark and provides contrast. Emulsions prepared under the following conditions: number of passes $N = 6$, homogenizer inlet air pressure $P = 130$ p.s.i., block copolypeptide concentration $C = 1.0$ mM, and oil volume fraction $\phi = 0.20$. **a**, Image of a WOW double emulsion stabilized by $K_{40}(\text{rac-L})_{20}$. **b**, Image of a single oil-in-water emulsion stabilized by $K_{60}L_{20}$. **c**, Image of size-fractionated droplets isolated from a $K_{40}(\text{rac-L})_{20}$ -stabilized double emulsion by low-speed centrifugation followed by ultracentrifugation. All scale bars, 200 nm.

Formation of nanoscale emulsion droplets is necessary for many applications, such as drug delivery where the outer droplet diameter generally needs to be less than 200 nm, and preferably between 50 nm and 100 nm (ref. 22). Although many methods are available for preparation of double emulsions, none allows preparation of outer droplets in this size range^{7–9,14}. We used ultrasonic homogenization to prepare a $K_{40}(\text{rac-L})_{20}$ emulsion yielding a polydisperse sample with the smallest double emulsion droplets observed by CTEM being around 400 nm in diameter. These droplets were further reduced in size by passage six times through a microfluidic homogenizer, yielding droplet diameters ranging from about ten to a few hundred nanometres. The stability of these double emulsions against both external and internal coalescence allowed the use of centrifugation to fractionate droplets into a desired size range. Centrifugation of the sample in Fig. 2a gave a buoyant fraction containing droplets hundreds of nanometres in diameter. The smaller droplets in the remaining suspension were further separated by ultracentrifugation¹¹, yielding a fraction with droplets ranging from about 10 to 100 nm in diameter (Fig. 2c). This fractionation procedure shows that isolation of stable double emulsion droplets in the nanoscale range is feasible, and that they are remarkably stable to shear.

To demonstrate their encapsulating ability, we loaded both water-soluble and oil-soluble fluorescent markers into copolypeptide-stabilized double emulsions. Water-soluble InGaP/ZnS quantum dots were mixed with fluorescein-labelled $K_{40}(\text{rac-L})_{10}$ before emulsification with silicone oil containing pyrene. Using fluorescence microscopy, we imaged both markers and the labelled polypeptide in the double emulsion droplets (Fig. 3a). The images also showed the compartmentalization of hydrophilic quantum dots (red) into the inner aqueous phase, hydrophobic pyrene (blue) into the oil phase and the labelled polypeptide (green) stabilizing the outer interface. Polypeptide at the inner interface was not observed, probably owing to quenching of the fluorescein label by the quantum dots. In samples prepared with $K_{60}L_{20}$ surfactant, we observed only simple oil droplets with no internal aqueous compartment (Fig. 3b). These cargos remained encapsulated within the droplets for at least three months, showing unprecedented stability of the inner aqueous compartment compared with other double emulsion systems^{4,7,14}.

Our $K_x(\text{rac-L})_y$ surfactants were designed with high hydrophilic contents, namely the ratio of hydrophilic to hydrophobic residues, which favour stabilization of oil-in-water emulsions where the oil is

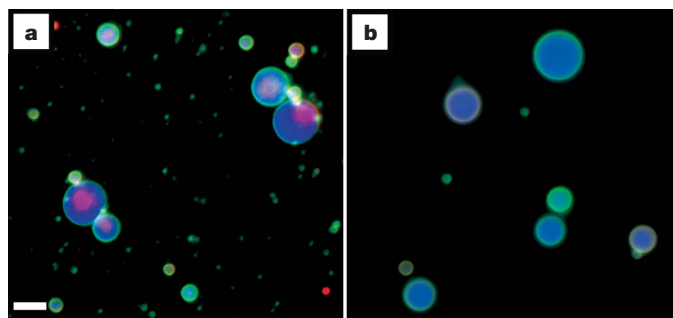


Figure 3 | Fluorescence micrographs of double emulsions containing polar and non-polar cargos. Samples were prepared using an ultrasonic homogenizer (10 s at 35% power) with $\phi = 0.2$ and $C = 0.1$ mM. The oil phase fluoresces blue because of entrapped pyrene (0.01 M), and the internal aqueous phase fluoresces red because of encapsulation of InGaP quantum dots (2 μ M). The polypeptides are labelled with fluorescein isothiocyanate (FITC) and therefore fluoresce green. Before imaging, the droplets were dialysed against and subsequently diluted with pure water to remove red fluorescence from the external phase (see Supplementary Information). **a**, Double emulsion stabilized by FITC-labelled $K_{40}(\text{rac-L})_{10}$, loaded with both pyrene and quantum dots. **b**, Single emulsion stabilized by FITC-labelled $K_{60}L_{20}$, loaded with both pyrene and quantum dots. Scale bars, 5 μ m.

on the concave side of the curved interface of a nanoscale droplet. Conversely, the inner water–oil interface of a WOW double emulsion is best stabilized by a surfactant with a low hydrophilic content because the oil is on the convex side of the interface. The opposite signs of these mean interfacial curvatures²³ explain why single-component surfactants generally do not stabilize double emulsion droplets and combinations of surfactants are required². This also explains the formation of only oil-in-water emulsions with $K_{60}L_{20}$, because the rod-like oligoleucine segments are poorly solvated by the oil and aggregate in the oil phase¹⁷. To stabilize an inner aqueous droplet in a WOW double emulsion, the hydrophobic polypeptide segments need to disperse in the oil to prevent steric crowding of the large hydrophilic segments in the aqueous phase (Fig. 1c).

The racemic-leucine segments in $K_x(rac-L)_y$ provide a combination of features that stabilize double emulsion droplets. The conformational flexibility of these segments improves oil solubility, because poly(*rac*-leucine) is soluble in organic solvents such as CH_2Cl_2 and $(CH_3)_2SO$ whereas poly(*L*-leucine) is not^{19,20}. This allows $K_x(rac-L)_y$ chains to stabilize the oil–water interface better in the inner droplet as the hydrophobic segments can disperse more readily in the oil. Despite its improved solubility, in an oil solvent nearly all residues of poly(*rac*-leucine) will also be engaged in both intramolecular and intermolecular hydrogen bonds. Studies on racemic polymers of both leucine and phenylalanine have demonstrated that they associate in organic solvents through hydrogen bonding²¹. At the interface of an inner aqueous droplet with oil, the high hydrophilic content of our polymers favours a low packing density of *rac*-leucine segments in the oil phase that would allow few interchain hydrogen bonds and give a weakly stabilized interface (Fig. 1c). But the opposite curvature of the oil–water interface in the outer droplet allows dense packing of the *rac*-leucine segments in the oil phase, favouring interchain hydrogen bonding. Consequently, even though inner aqueous droplets are likely to be unstable, they are prevented from merging with the outer droplets, and forming simple emulsions, as the outer interfaces are expected to be reinforced by hydrogen-bond cross-linking. To test this concept, emulsions were prepared containing a silicone oil capped with acetamide groups capable of hydrogen bonding to *rac*-leucine segments. Emulsification with $K_{60}(rac-L)_{20}$ gave WOW nanoemulsions containing multiple internal droplets (Supplementary Fig. 6), supporting the hypothesis that *rac*-leucine segments can stabilize droplets through hydrogen bonding interactions in the oil phase, thus inhibiting internal droplet coalescence.

Our use of racemic, disordered hydrophobic polypeptide segments that interact through hydrogen bonding is a new means of stabilizing WOW double emulsions. This approach differs greatly from protein- and peptide-stabilized emulsions where double emulsions do not form without the use of additional surfactants, and an ordered amphiphilic helix is the most common source of surface activity^{24–28}. Our strategy also can be applied to other copolypeptides, because samples containing *rac*-valine and *rac*-alanine hydrophobic segments also gave stable double nanoemulsions (Supplementary Fig. 3b,c). Use of block copolypeptide surfactants overcomes key limitations of WOW double emulsions by allowing the straightforward preparation of stable nanoscale droplets that can simultaneously encapsulate both oil-soluble and water-soluble cargos.

METHODS SUMMARY

We first dissolved $K_{40}(rac-L)_{20}$ copolypeptide in ultrapure water at the desired concentration (0.01–1.5 mM). Silicone oil (viscosity $0.1\text{ cm}^2\text{ s}^{-1}$) was added to give the desired volume fraction (ϕ) of oil in the continuous phase ($0.05 < \phi < 0.50$). We prepared a microscale emulsion either by mixing for 1 min using a hand-held homogenizer (IKA Ultra-Turrax T8 with the S8N-8G dispersing element) or by mixing for 10 s using a hand-held ultrasonic homogenizer (Cole-Parmer 4710 Series Model ASI at an output of 35–40%). This emulsion was then passed through a processor (M-110S Microfluidizer) with a 75- μm stainless steel/ceramic interaction chamber and an input air pressure (P) of 130 p.s.i. The emulsion was collected at the product outlet, and then passed through the microfluidic homogenizer repeatedly for a total of six passes

($N = 6$), which decreased the average droplet radius ($\langle a \rangle$) and increased the monodispersity of the sample. We used a similar protocol for emulsions generated using other block copolypeptide surfactants (Supplementary Table 1, Supplementary Fig. 2a–c). The ratio of inner droplet radius to outer droplet radius was relatively uniform for different hydrophobic chain lengths at about 0.5 (Supplementary Table 1, Supplementary Fig. 2d). Other amphiphilic block copolypeptides where either the lysine or leucine domains were substituted with different hydrophilic or hydrophobic residues, respectively, also formed double emulsions (Supplementary Fig. 3a–d). We also qualitatively evaluated the emulsification capability of different polypeptide surfactants using toluene, which forms less stable emulsions, and with a control homopolypeptide, K_{60} (Supplementary Fig. 4).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 December 2007; accepted 23 June 2008.

1. Bibette, J., Calderon, F. L. & Poulin, P. Emulsions: Basic principles. *Rep. Prog. Phys.* **62**, 969–1033 (1999).
2. Ficheux, M. F., Bonakdar, L., Leal-Calderon, F. & Bibette, J. Some stability criteria for double emulsions. *Langmuir* **14**, 2702–2706 (1998).
3. Wang, Y. F., Tao, Z. & Gang, H. Structural evolution of polymer-stabilized double emulsions. *Langmuir* **22**, 67–73 (2006).
4. Pays, K. *et al.* Double emulsions: how does release occur? *J. Control. Release* **79**, 193–205 (2002).
5. Davis, S. S. & Walker, I. M. Multiple emulsions as targetable delivery systems. *Methods Enzymol.* **149**, 51–64 (1987).
6. Okochi, H. & Nakano, M. Preparation and evaluation of W/O/W type emulsions containing vancomycin. *Adv. Drug Deliv. Rev.* **45**, 5–26 (2000).
7. Garti, N. Double emulsions: Scope, limitations and new achievements. *Colloids Surf. A* **123**, 233–246 (1997).
8. Loscertales, I. G. *et al.* Micro/nano encapsulation via electrified coaxial liquid jets. *Science* **295**, 1695–1698 (2002).
9. Utada, A. S. *et al.* Monodisperse double emulsions generated from a microcapillary device. *Science* **308**, 537–541 (2005).
10. Morais, J. M., Santos, O. D. H., Nunes, J. R. L., Zanatta, C. F. & Rocha-Filho, P. A. W/O/W multiple emulsions obtained by one-step emulsification method and evaluation of the involved variables. *J. Disp. Sci. Technol.* **29**, 63–69 (2008).
11. Mason, T. G., Wilking, J. N., Meleson, K., Chang, C. B. & Graves, S. M. Nanoemulsions: formation, structure, and physical properties. *J. Phys. Condens. Matter* **18**, R635–R666 (2006).
12. Goubault, C. *et al.* Shear rupturing of complex fluids: Application to the preparation of quasi-monodisperse water-in-oil-in-water double emulsions. *Langmuir* **17**, 5184–5188 (2001).
13. Okushima, S., Nisisako, T., Torii, T. & Higuchi, T. Controlled production of monodisperse double emulsions by two-step droplet breakup in microfluidic devices. *Langmuir* **20**, 9905–9908 (2004).
14. Benichou, A., Aserin, A. & Garti, N. Double emulsions stabilized with hybrids of natural polymers for entrapment and slow release of active matters. *Adv. Colloid Interface Sci.* **108–109**, 29–41 (2004).
15. Katchalski, E. & Sela, M. Synthesis and chemical properties of poly- α -amino acids. *Adv. Protein Chem.* **13**, 243–492 (1958).
16. Niederhafner, P., Šebestík, J. & Ježek, J. Peptide dendrimers. *J. Pept. Sci.* **11**, 757–788 (2005).
17. Nowak, A. P. *et al.* Rapidly recovering hydrogel scaffolds from self-assembling diblock copolypeptide amphiphiles. *Nature* **417**, 424–428 (2002).
18. Holowka, E. P., Pochan, D. J. & Deming, T. J. Charged polypeptide vesicles with controllable diameter. *J. Am. Chem. Soc.* **127**, 12423–12428 (2005).
19. Kricheldorf, H. R. & Mang, T. C-13-NMR sequence-analysis, 20. Stereospecificity of the polymerization of D,L-Leu-NCA and D,L-Val-NCA. *Makromol. Chem.: Macromol. Chem. Phys.* **182**, 3077–3098 (1981).
20. Breitenbach, J. W., Allinger, K. & Koref, A. Viskositätsstudien an Lösungen von DL-Phenylalanin-Polypeptiden. *Monatsh. Chem.* **86**, 269 (1955).
21. Lapp, C. & Marchal, J. Preparation de la poly-D,L-phenylalanine en helice par polymerisation de la D,L-benzyl-4 oxazolidine dione-2-5. *J. Chim. Phys.* **60**, 756–766 (1963).
22. Kataoka, K., Kwon, G. S., Yokoyama, M., Okano, T. & Sakurai, Y. Block-copolymer micelles as vehicles for drug delivery. *J. Control. Release* **24**, 119–132 (1993).
23. Strey, R. Microemulsion microstructure and interfacial curvature. *Colloid Polym. Sci.* **272**, 1005–1019 (1994).
24. Enser, M., Bloomberg, G. B., Brock, C. & Clark, D. C. De novo design and structure–activity relationships of peptide emulsifiers and foaming agents. *Int. J. Biol. Macromol.* **12**, 118–124 (1990).
25. Dickinson, E. Structure and composition of adsorbed protein layers and the relationship to emulsion stability. *J. Chem. Soc., Faraday Trans.* **88**, 2973–2983 (1992).
26. Saito, M., Ogasawara, M., Chikuni, K. & Shimizu, M. Synthesis of a peptide emulsifier with an amphiphilic structure. *Biosci. Biotechnol. Biochem.* **59**, 388–392 (1995).

27. Dalgleish, D. G. Conformations and structures of milk proteins adsorbed to oil–water interfaces. *Food Res. Intl* **29**, 541–547 (1996).
28. Chang, C. B., Knobler, C. M., Gelbart, W. M. & Mason, T. G. Curvature dependence of viral protein structures on encapsidated nanoemulsion droplets. *ACS Nano* **2**, 281–286 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is supported by a grant from the National Science Foundation (T.J.D.), a grant from the Human Frontiers of Science Program (T.J.D.), and University of California start-up funds (T.J.D. and T.G.M.). Some of the work was conducted at the National Resource for Automated Molecular Microscopy (NRAMM) which is supported by the National Institutes of Health through the National Center for Research Resources P41 programme. We thank C. Potter and

J. Quispe of NRAMM at the Scripps Research Institute, and S. Zhong and D. Pochan of the University of Delaware, for use of cryoelectron microscopy equipment.

Author Contributions J.A.H. synthesized the polypeptides, prepared and characterized the emulsions, designed experiments, and assisted in manuscript preparation. C.B.C and S.M.G. also prepared and characterized the emulsions. Z.L. imaged the emulsions using CTEM. T.G.M. and T.J.D. initiated the project, designed and supervised the experiments, analysed the data, and drafted the manuscript. T.J.D. assembled and finalized the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to T.J.D. (demingt@seas.ucla.edu) or T.G.M. (mason@chem.ucla.edu).

METHODS

Fractionation of emulsions. A $K_{40}(\text{rac-L})_{20}$ emulsion, with block copolypeptide concentration $C = 1.5 \text{ mM}$ (prepared as in Methods Summary), was centrifuged in a 15 mL plastic centrifuge tube for 24 h at 3,500 r.p.m. using a tabletop centrifuge (IEC HN-S). A 0.5-mm plug was formed and separated from the remnant suspension beneath. The plug formed at the top of the tube because the density of silicone oil is lower than water (0.973 g mL^{-1} for $0.1 \text{ cm}^2 \text{ s}^{-1}$ silicone oil, 1.0 g mL^{-1} for water). The remnant suspension was further fractionated at 20,000 r.p.m. for 4 h using an ultracentrifuge (Beckman L8-55) with a swinging bucket rotor. The plug that formed on top of the suspension was separated and the remaining suspension was imaged using CTEM (Fig. 2c).

Dynamic light scattering. Because the interfacial organization of double emulsions is complex, describing their structure in complete detail can be complicated. Two different droplet size distributions are necessary for inner and outer droplets, $p_i(a_i)$ and $p_o(a_o)$, respectively, where a is the radius. Although the droplet volume fraction of the outer droplets is simply ϕ_o , the distribution of inner droplet volume fractions depends on $p_i(a_i)$ and on the number distribution of smaller droplets within a given droplet, $p_i(N_i)$, where N_i is the number of inner droplets. To simplify the description of double emulsions, usually average radii (for example, \bar{a}_i and \bar{a}_o), inner volume fractions ϕ_i and numbers of inner droplets \bar{N}_i are reported, as quantifying the full distributions can be difficult. The outer diameters of emulsion droplets were estimated by dynamic light scattering (DLS) with a Photocor-FC board and software. Although DLS of double emulsions yields intensity correlation decay data that are complex³, we believe the DLS data provide a crude estimate of average outer droplet diameter consistent with CTEM real-space data. Average outer droplet diameters from CTEM measurements were generally lower than diameters from DLS, reflecting the inevitable exclusion of larger droplets from the thin vitrified water layer ($<200 \text{ nm}$) usable for CTEM imaging. The DLS samples were diluted to obtain an intensity reading of between 1×10^5 and 6×10^5 counts. Each sample was run at 90° scattering angle for 500 s, with linear channel spacing and an adjustable baseline. The fitting procedure used was cumulant analysis with an adjustable baseline to fit the data and calculate droplet radii. DLS data for different emulsion formulations are given in Supplementary Fig. 5.

Fluorescence microscopy. Before fluorescence imaging, emulsion suspensions were diluted tenfold with deionized water. A drop of emulsion was then placed on a glass slide and covered using a glass cover slip. The samples were imaged using a Zeiss Axiovert 200 fluorescence microscope equipped with ultraviolet filter set #49 (excitation 365 nm, emission 420 to 470 nm), blue filter set #10 (excitation 450 to 490 nm, emission 515 to 565 nm), and green filter set #43 (excitation 530 to 560 nm, emission 570 to 640 nm).

CTEM imaging. Each emulsion sample was diluted tenfold with deionized water before imaging. An aliquot of each sample ($5 \mu\text{L}$) was then placed on a Formvar stabilized with carbon 300 mesh copper grid (Ted Pella). The grid was loaded into a Vitrobot (FEI) automated vitrification device for automated sample blotting and vitrification in liquid ethane. The grid was stored under liquid nitrogen and then placed, using a cold stage, in a Phillips Tecnai F20 electron microscope and imaged with an accelerating voltage of 120 kV. Images were obtained on a Teitz SCX slow-scan CCD detector controlled by the Legimon software package²⁹.

Critical aggregation concentration via pyrene fluorescence. Polypeptide solutions (2 mL) were dispersed in water at a range of concentrations (2.0×10^{-3} to $2.0 \times 10^{-12} \text{ M}$). A stock pyrene solution was made by dissolving pyrene in acetone ($6.0 \times 10^{-2} \text{ M}$). Next, an appropriate amount of the pyrene stock solution was added to give a final concentration of $12 \times 10^{-7} \text{ M}$ in water and the acetone was evaporated off. To each polypeptide solution, we added 2.0 mL of the stock pyrene solution to afford a final concentration of $6.0 \times 10^{-7} \text{ M}$. Each solution was allowed to equilibrate overnight before measurements. To record fluorescence spectra, we added 3.0 mL of each polypeptide solution to a polystyrene cuvet (4.0 mL). The excitation spectra were recorded within a range of 300–360 nm at an emission wavelength of 390 nm. All spectra were run with an integration time of 1 s per 0.5 nm. The ratio of the intensities of two peaks I_{338}/I_{333} was plotted as a function of polypeptide concentration (M) for each sample. The critical aggregation concentrations were determined as the intersection of the extrapolated straight line fits of the plot as previously described¹⁸.

Interfacial tension measurements. Interfacial tension (γ) values between polypeptide solutions ($0.1 \text{ mM } K_{60}L_{20}$ and $0.1 \text{ mM } K_{40}(\text{rac-L})_{20}$) and PDMS ($0.1 \text{ cm}^2 \text{ s}^{-1}$) were measured using the Du Nouy ring method outlined by Zuidema and Waters³⁰. A platinum–iridium ring (circumference 5.0 cm) was attached to a balance and the mass of the oil/polypeptide solution interface was measured as the ring was pulled at a rate of 0.01 mm s^{-1} using a calibrated bottom-hole balance apparatus at 25°C . The polypeptide solutions ($K_{60}L_{20}$ and $K_{40}(\text{rac-L})_{20}$) were well above their measured critical aggregation concentration values of 7.1×10^{-7} and $9.7 \times 10^{-7} \text{ M}$, respectively. To reduce wall effects, the diameter of the container (8.0 cm) was significantly larger than the diameter of the Du Nouy ring. In addition, each polypeptide solution was equilibrated with the oil–water interface for at least 24 h before measurement.

29. Carragher, B. et al. Legimon: An automated system for acquisition of images from vitreous ice specimens. *J. Struct. Biol.* **132**, 33–45 (2000).

30. Zuidema, H. H. & Waters, G. W. Ring method for the determination of interfacial tension. *Ind. Eng. Chem. Anal. Edn* **13**, 312–313 (1941).

Interaction between liquid water and hydroxide revealed by core-hole de-excitation

Emad F. Aziz¹, Niklas Ottosson^{1,2}, Manfred Faubel³, Ingolf V. Hertel^{4,5} & Bernd Winter^{4,†}

The hydroxide ion plays an important role in many chemical and biochemical processes in aqueous solution¹. But our molecular-level understanding of its unusual and fast transport in water, and of the solvation patterns that allow fast transport, is far from complete. One proposal seeks to explain the properties and behaviour of the hydroxide ion by essentially regarding it as a water molecule that is missing a proton², and by inferring transport mechanisms and hydration structures from those of the excess proton. A competing proposal invokes instead unique and interchanging hydroxide hydration complexes, particularly the hypercoordinated $\text{OH}^-(\text{H}_2\text{O})_4$ species and tri-coordinated $\text{OH}^-(\text{H}_2\text{O})_3$ that can form a transient hydrogen bond between the H atom of the OH^- and a neighbouring water molecule^{3–5}. Here we report measurements of core-level photoelectron emission and intermolecular Coulombic decay^{6–8} for an aqueous hydroxide solution, which show that the hydrated hydroxide ion is capable of transiently donating a hydrogen bond to surrounding water molecules. In agreement with recent experimental studies of hydroxide solutions^{9–12}, our finding thus supports the notion that the hydration structure of the hydroxide ion cannot be inferred from that of the hydrated excess proton.

Core-level electron spectroscopy techniques probe the local electronic structure of molecules, which in turn provides indirect information about local structural details. During core-level electron spectroscopy measurements on aqueous hydroxide solutions, we have discovered the occurrence of intermolecular Coulombic decay (ICD), which involves coupled changes of the electronic structure of an OH^- ion and of a neighbouring water molecule. Because of the unique way in which the ICD process is connected with structure, it provides fairly direct information about local structural details involving the interacting OH^- ion and water molecule. The ICD measurements are combined with photoelectron spectroscopy to determine absolute electron energies of OH^- in liquid water and allow a robust assignment to ICD. Both techniques necessitate the non-trivial task of measuring electron kinetic energies for highly volatile solutions.

Our electron emission measurements were made on a 15 μm liquid jet of a 4 molal NaOH aqueous solution at 4 °C, using undulator synchrotron radiation from the U41-PGM beamline at BESSY, Berlin. Experimental details are as previously described^{13,14}. Figure 1 shows in red the four electron spectra obtained with excitation photon energies that fall near or in the X-ray absorption band of $\text{OH}^-(\text{aq})$ (the 'A-band') around 532.8 eV (ref. 12). (The A-band is clearly apparent in the oxygen K-edge X-ray absorption spectrum, shown in Supplementary Fig. 1b.) For comparison, the corresponding spectra of neat liquid water are shown in blue. The excitation photon energy of 531.0 eV giving rise to trace a is just below the onset energy of the A-band, so the spectra in Fig. 1a arise from direct photoelectron emission. The difference between the solution and pure water spectra,

at electron kinetic energies of 495.6 and 521.8 eV, is due to photoemission from $\text{Na}^+(\text{aq})$ 2p and $\text{OH}^-(\text{aq})$, respectively; these values correspond to electron binding energies of 35.4 and 9.2 eV. Spectra measured for excitation at the high-energy tail of the $\text{OH}^-(\text{aq})$ A-band, using photons with energy of 534.0 eV, are shown in trace d. The solution and water spectra are again rather similar, with a broad emission band, but differ strongly from the corresponding spectra in Fig. 1a. The dominant band is due to normal Auger-electron emission from liquid water¹⁴, with less than 20% of the spectral intensities originating from direct photoelectrons. The small feature 1 at

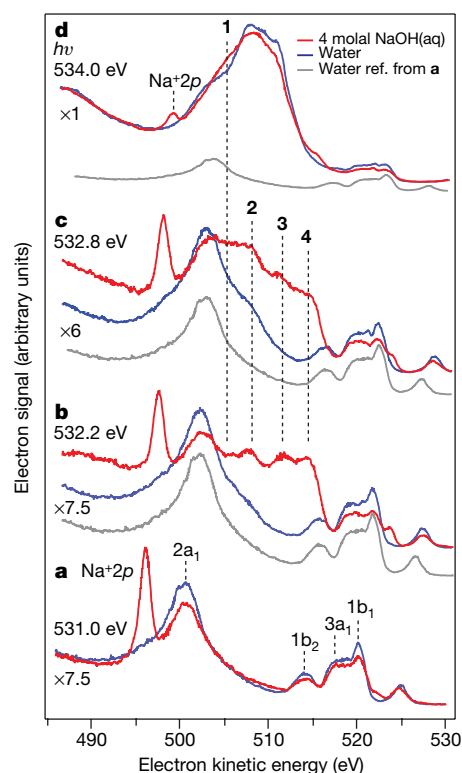


Figure 1 | Photoelectron, resonant Auger-electron and intermolecular Coulombic decay spectra of 4 molal NaOH aqueous solution. Spectra of neat liquid water are shown for comparison. **a**, Excitation at the low-energy onset of the A-band; **d**, the high-energy tail. **b**, **c**, Spectra measured at energies slightly lower and higher, respectively, than the A-band maximum. Resonance peaks **2–4** reveal the existence of the OH^- hydrogen donor bond. The grey spectrum is the relative contribution from direct photoemission. The small peak at highest kinetic energies arises from ionization at $2h\nu$, providing a very exact energy calibration.

¹BESSY GmbH, Albert-Einstein-Strasse 15, D-12489 Berlin, Germany. ²Department of Physics, Uppsala University, SE-75121 Uppsala, Sweden. ³Max-Planck-Institut für Dynamik und Selbstorganisation, Bunsenstrasse 10, D-37073 Göttingen, Germany. ⁴Max-Born-Institut für Nichtlineare Optik und Kurzzeitspektroskopie, Max-Born-Strasse 2a, D-12489 Berlin, Germany. ⁵Institut für Experimentalphysik, Freie Universität Berlin, Arnimallee 14, D-14195 Berlin, Germany. [†]Present address: BESSY GmbH, Albert-Einstein-Strasse 15, D-12489 Berlin, Germany.

505.3 eV kinetic energy in the NaOH(aq) spectrum is due to normal Auger-electron emission following $\text{OH}^-(\text{aq})$ oxygen 1s ionization, which is possible at this threshold energy. As we now tune the excitation energy to the absorption maximum of the A-band (Fig. 1b, c), the dominant Auger-electron contribution from water disappears and the water spectra barely differ from the undisturbed photoemission spectrum in trace a. These changes in the water spectrum allow us to detect new resonance features that appear in the solution spectra, with peaks 2–4 at 508.2, 512.0 and 514.3 eV kinetic energy clearly visible against a smooth background. Equally well-resolved resonance spectral structures are observed for KOH aqueous solution (data not shown).

The resonance peaks 2–4 in the solution spectra in Fig. 1 appear at kinetic energies that are too high to originate from resonance Auger-electron emission, even if the emission involved a shakedown process^{14,15}. A hint as to their origin comes from the observation that the kinetic energies of 2–4, at 508.2, 512.0 and 514.3 eV, show exactly the same energy spacings as the three outer valence orbitals of a water molecule. Moreover, the absolute kinetic energies are almost identical to the kinetic energy values of the emission peaks obtained in photoelectron spectroscopy measurements of neat water when using 526.8 eV photons. The energy of 526.8 eV appears to correspond to the electronic energy released on refilling the oxygen 1s core hole of $\text{OH}^-(\text{aq})$ by a $\text{OH}^-(\text{aq})$ $2p\pi$ valence electron, as illustrated by transition C in the energy-level diagram of Fig. 2c. Taken together, these considerations suggest that the excess energy released when filling the core hole of $\text{OH}^-(\text{aq})$ is fully transferred from $\text{OH}^-(\text{aq})$ to a neighbouring water molecule, ionizing the latter's outer valence orbitals $1b_2$, $3a_1$ and $1b_1$. By using the photoelectric law, kinetic energy $\text{KE} = h\nu - \text{BE}$ (where BE is binding energy) and the electron binding energies of the relevant $\text{OH}^-(\text{aq})$ and water molecule levels (see Fig. 2), the

photoelectron kinetic energies associated with ionizing $1b_2$, $3a_1$ and $1b_1$ can be estimated. The values obtained are 509.5, 513.3 and 515.6 eV, respectively, which are in good agreement with experiment.

Our measurements thus document intermolecular Coulombic decay involving OH^- in an aqueous solution. ICD, the transfer of energy from an atom or molecule to a neighbouring atom or molecule and subsequent ionization of the latter, is ubiquitous in weakly bonded systems and has been observed in van der Waals¹⁶ and hydrogen-bonded clusters, including small water clusters (U. Hergenbahn, personal communication). But the present work is the first to detect the process in aqueous solution, and to initiate it through core-hole excitation (which gives rise to initial excited states with very short lifetimes that are determined by the lifetime of the core hole, and are approximately 4 fs for the oxygen 1s core-level). The present observation contrasts with the results of our recent studies of $2p$ edge excitation of chloride in aqueous solution¹⁵ (and also of 1s excitation of fluoride in aqueous solution; see Supplementary Fig. 3), where we identified a manifold of charge-transfer-to-solvent (CTTS) states and associated electron dynamics through the occurrence of spectator Auger-electron peaks. CTTS states are possible when after excitation, the electron is immediately bound in a potential well that arises because of the pre-existing polarization of oriented solvent dipoles around the ion¹⁷. This phenomenon may be qualitatively different for halide anions¹⁸ than for OH^- (ref. 19), given the differences between their hydration patterns. Still, whereas spectator peaks in aqueous chloride solution spectra identify resonances and signal the localization of the CTTS electron on the subfemtosecond timescale of the core hole (see ref. 15 and Supplementary Figs 2 and 3), we find in this work that resonant excitation of $\text{OH}^-(\text{aq})$ does not give rise to local Auger decay even though the excited states of $\text{OH}^-(\text{aq})$ are of CTTS nature²⁰. Instead, a new and competitive (on an ultrafast timescale) relaxation mechanism, ICD, opens up.

ICD requires favourable orientations and distances between the donor and acceptor molecules, to allow for sufficient orbital overlap yet only negligible rehybridization. As opposed to the local Auger decay, for ICD to occur it is sufficient that the orbital of the initial core-hole vacancy overlaps with the orbital of the resulting final hole at the same molecular site; the second hole has delocalized to a neighbouring site⁶. Our present results, in combination with the earlier observations on aqueous halide solutions, suggest that ICD rates can change greatly with relatively subtle changes in hydration pattern. Specifically, we attribute the fact that ICD is observed only in the present study to the ability of $\text{OH}^-(\text{aq})$ to donate a hydrogen bond to a neighbouring water molecule, along which ICD can occur. This selectivity might be directly associated with the directional lobe for the electron localization function at the hydrogen site of the hydroxide ion, which contrasts with a ring-like structure⁵ for the electron localization function at the ion's oxygen site. In support of such OH^- donor-H-bond specificity in ICD, we note that no characteristic ICD pattern (that is, mirroring ionization of the water valence orbitals) is found in the halide spectra, not even in the spectrum of $\text{F}^-(\text{aq})$, which is isoelectronic to OH^- . Both anions have almost the same final-state solvation energy, they have similar electronegativity and form hydrogen bonds with charge-transfer character that involve the hydrogen atoms of water molecules, and their ionic radii are almost the same (1.33 Å and 1.32 Å). Given all these similarities, we conclude that the unique resonance spectral features in the hydroxide spectra of Fig. 1 must arise because of the presence of this anion's extra hydrogen atom.

A direct and model-independent consequence of this interpretation of the unique spectral features is that it calls for a $\text{OH}^-(\text{aq})$ hydration pattern similar to the one invoked^{4,5} in one of the two competing mechanisms put forward to explain the anomalously fast transport of OH^- in aqueous solution. This mechanism assumes that the oxygen in OH^- is on average hypercoordinated and preferentially accepting four hydrogen bonds, forming the complex $\text{OH}^-(\text{H}_2\text{O})_4$; the key to OH^- transport is the conversion into a tetrahedral tri-coordinated complex $\text{OH}^-(\text{H}_2\text{O})_3$ by proton transfer, and subsequent formation of a transient OH^- hydrogen-donor bond to give a $\text{OH}^-(\text{aq})$

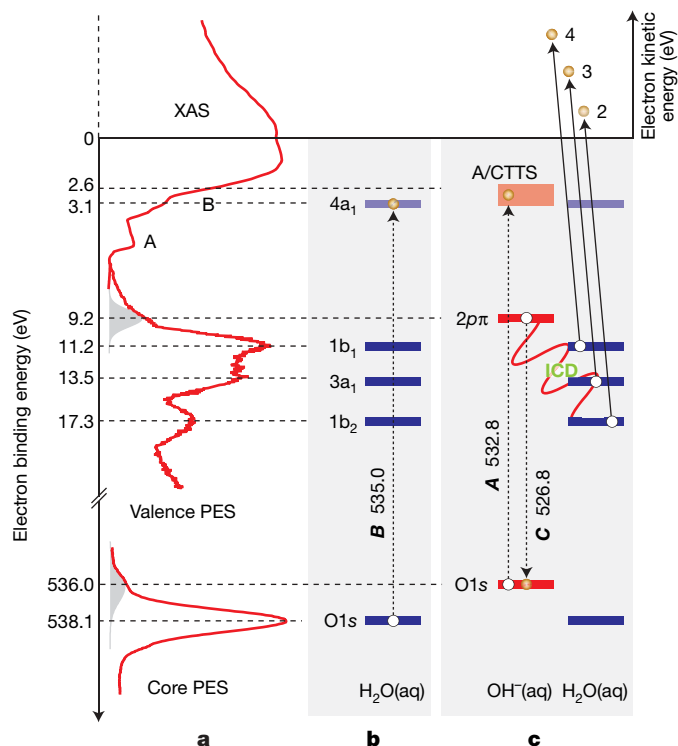


Figure 2 | Energy-level diagram of $\text{OH}^-(\text{aq})$ and $\text{H}_2\text{O}(\text{aq})$. **a**, Experimental spectra of a 4 molal NaOH aqueous solution. The $\text{H}_2\text{O}(\text{aq})$ and $\text{OH}^-(\text{aq})$ oxygen 1s photoemission spectra (PES) reveal binding energies of 538.1 and 536.0 eV (grey peak), respectively. The X-ray absorption spectrum (XAS) of $\text{OH}^-(\text{aq})$ has a maximum at 532.8 eV. **b**, $\text{H}_2\text{O}(\text{aq})$ oxygen 1s $\rightarrow 4a_1$ resonant excitation at 535.0 eV photon energy. The absolute energy position of the X-ray absorption spectrum is with respect to water rather than to $\text{OH}^-(\text{aq})$. Hence the water B-band position of the spectrum coincides with the $4a_1$ energy, and the $\text{OH}^-(\text{aq})$ A-band is not aligned with the CTTS states in the figure. **c**, Illustration of ICD for $\text{OH}^-(\text{aq})$.

hydration topology resembling that of a water molecule^{4,5}. The alternative transport mechanism is based on the 'proton-hole' concept^{2,21}, which treats the hydroxide ion as a water molecule with a missing proton and OH[−] transport as the mirror image process of proton structural diffusion. In this picture, tri-coordinated OH[−](H₂O)₃ is predicted² to be the dominant OH[−] hydration pattern and OH[−](aq) assumed to be unable to donate a hydrogen bond (in analogy with the inability of the Eigen proton complex, H₃O⁺(H₂O)₃, to accept a hydrogen bond via its lone pair). Our findings are not in agreement with the predictions of the proton-hole concept. In fact, neutron diffraction^{10,11} and Fourier transform infrared spectroscopy²² investigations have also concluded that OH[−](H₂O)₄ is the dominant OH[−] hydration pattern in aqueous solution. The neutron diffraction studies furthermore revealed the presence of a (probably weakly hydrogen-bonded) water molecule near the OH[−] hydrogen atom, in agreement with the results of a combined X-ray diffraction and simulation study¹². We note that the overall conclusion of an asymmetry in the hydration patterns for the proton and hydroxide ion may affect not only discussions of proton and hydroxide transport in aqueous solutions, but also our understanding of interfacial solvation behaviour and the question of whether the surface of liquid water is enriched in either hydrated protons or hydroxide ions^{23–25}. Protons are known to disrupt the hydrogen bonding in liquid water and are preferentially accommodated at surfaces²⁶. Although our experiments probe beyond the liquid surface (covering an integrated depth of about 20 Å), the fundamental asymmetry between H₃O⁺(aq) (unable to accept a hydrogen bond) and OH[−](aq) (capable of hydrogen-bond donation) suggests that a basic aqueous solution surface is not rich in hydroxide.

As noted above, a plethora of experimental data supports the presence of different coexisting OH[−] hydration structures in aqueous solution as demanded by the first of the OH[−] transport mechanisms discussed. Coexisting structures, with different stabilization energies, are also the likely reason for the observed width of the OH[−](aq) absorption band (note the width of the A/CTTS band in Fig. 2c). In contrast, ICD selectively accesses the part of the A-state manifold that corresponds to short-lived OH[−](aq) hydration patterns with donated hydrogen bonds that aid fast ICD. Because the width of the A-band is given by the full ensemble of hydration structures, with the OH[−](H₂O)₄ acceptor-only configurations as well as the transient OH[−](H₂O)₃ patterns being the crucial limiting structures, we expect the excitation range resulting in ICD to be narrower than the full A-band. This is indeed observed experimentally, and directly confirms that hydration of OH[−] involves different types of hydrogen bonds. No such distinction exists for the halide anions, which are hydrogen-bond acceptors only. We conclude that the resonance spectral structure observed in this study must be a microscopic signature of the OH[−](aq) hydrogen-donor bond. Decomposition of the A-state manifold into contributions from different OH[−](aq) hydration patterns could in principle reveal the lifetime of the transient hydrogen donor structure, but configurational broadening makes it difficult to extract a reliable value from the spectra.

We conclude by noting that our observation of a transient hydrogen bond donated by OH[−], in conjunction with the hypercoordinated structures discerned in neutron and X-ray diffraction studies of macroscopic OH[−] aqueous solution^{10–12}, suggests a hydroxide solvation behaviour distinctively different from that inferred from spectroscopic studies on gas-phase OH[−](H₂O)_n clusters where the weak hydrogen-donor bond is unable to form²⁷. This difference highlights the importance of long-range water-solvent behaviour, and also the need for sophisticated experiments in the bulk liquid phase to help formulate and test detailed descriptions of bulk aqueous solution properties. X-ray absorption spectroscopy has, for instance, revealed⁹ the distinct OH[−](aq) X-ray absorption band at 532.8 eV (see Supplementary Fig. 1b) that serves as an exclusive fingerprint of the hydroxide anion itself, favourably assuming the hypercoordinated pattern. The latter was inferred from the fundamentally different interaction between OH[−] and solvating water molecules and the corresponding interactions of aqueous halide anions²⁸. The present electron-energy spectra measured at the oxygen

1s resonance provide valuable complementary information, in that they allow a more direct identification of the hypercoordinated pattern and also allow the transient hydrogen bond donated by OH[−] to be captured.

Received 30 November 2007; accepted 9 July 2008.

- Hynes, J. T., Klinman, J. P., Limbach, H.-H. & Schowen, R. L. E. *Hydrogen-Transfer Reactions* Ch. 21 (Wiley, 2007).
- Asthagiri, D., Pratt, L. R., Kress, J. D. & Gomez, M. A. Hydration and mobility of HO[−](aq). *Proc. Natl Acad. Sci. USA* **101**, 7229–7233 (2004).
- Chandra, A., Tuckerman, M. E. & Marx, D. Connecting solvation shell structure to proton transport kinetics in hydrogen-bonded networks via population correlation functions. *Phys. Rev. Lett.* **99**, 145901 (2007).
- Tuckerman, M. E., Chandra, A. & Marx, D. Structure and dynamics of OH[−](aq). *Acc. Chem. Res.* **39**, 151–158 (2006).
- Tuckerman, M. E., Marx, D. & Parrinello, M. The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature* **417**, 925–929 (2002).
- Cederbaum, L. S., Zobeley, J. & Tarantelli, F. Giant intermolecular decay and fragmentation of clusters. *Phys. Rev. Lett.* **79**, 4778–4781 (1997).
- Averbukh, V., Müller, I. B. & Cederbaum, L. S. Mechanism of interatomic Coulombic decay in clusters. *Phys. Rev. Lett.* **93**, 263002 (2004).
- Müller, I. B. & Cederbaum, L. S. Ionization and double ionization of small water clusters. *J. Chem. Phys.* **125**, 204305 (2006).
- Cappa, C. D., Smith, J. D., Messer, B. M., Cohen, R. C. & Saykally, R. J. Nature of the aqueous hydroxide ion probed by X-ray absorption spectroscopy. *J. Phys. Chem. A* **111**, 4776–4785 (2007).
- Botti, A., Bruni, F., Imberti, S., Ricci, M. A. & Soper, A. K. Ions in water: The microscopic structure of concentrated NaOH solutions. *J. Chem. Phys.* **120**, 10154–10162 (2004).
- Imberti, S. *et al.* Ions in water: The microscopic structure of concentrated hydroxide solutions. *J. Chem. Phys.* **122**, 194509 (2005).
- Megyes, T. *et al.* The structure of aqueous sodium hydroxide solutions: A combined solution X-ray diffraction and simulation study. *J. Chem. Phys.* **128**, 044501–044512 (2008).
- Winter, B. & Faubel, M. Photoemission from liquid aqueous solutions. *Chem. Rev.* **106**, 1176–1211 (2006).
- Winter, B., Hergenroth, U., Faubel, M., Björneholm, O. & Hertel, I. V. Hydrogen bonding in liquid water probed by resonant Auger-electron spectroscopy. *J. Chem. Phys.* **127**, 094501 (2007).
- Winter, B. *et al.* Electron dynamics in charge-transfer-to-solvent states of aqueous chloride revealed by Cl[−] 2p resonant Auger-electron spectroscopy. *J. Am. Chem. Soc.* **130**, 7130–7138 (2008).
- Barth, S. *et al.* Interface identification by non-local autoionization transitions. *Phys. Chem. Chem. Phys.* **8**, 3218–3222 (2006).
- Chen, X. & Bradforth, S. E. The ultrafast dynamics of photodetachment. *Annu. Rev. Phys. Chem.* **59**, 203–231 (2008).
- Blandamer, M. J. & Fox, M. F. Theory and applications of charge-transfer-to-solvent spectra. *Chem. Rev.* **70**, 59–93 (1970).
- Crowell, R. A. *et al.* Ultrafast dynamics for electron photodetachment from aqueous hydroxide. *J. Chem. Phys.* **120**, 11712–11725 (2004).
- Fox, M. & McIntyre, R. Far ultraviolet solution spectroscopy of hydroxide. *Faraday Discuss.* **64**, 167–172 (1977).
- Agmon, N. Mechanism of hydroxide mobility. *Chem. Phys. Lett.* **319**, 247–252 (2000).
- Smiechowski, M. & Stangret, J. Hydroxide ion hydration in aqueous solutions. *J. Phys. Chem. A* **111**, 2889–2897 (2007).
- Vacha, R., Buch, V., Milet, A., Devlin, J. P. & Jungwirth, P. Autoionization at the surface of neat water: Is the top layer pH neutral, basic, or acidic? *Phys. Chem. Chem. Phys.* **9**, 4736–4747 (2007).
- Beattie, J. K. Comment on Autoionization at the surface of neat water: is the top layer pH neutral, basic, or acidic? by R. Vacha, V. Buch, A. Milet, J. P. Devlin and P. Jungwirth. *Phys. Chem. Chem. Phys.*, 2007, **9**, 4736. *Phys. Chem. Chem. Phys.* **10**, 330–331 (2008).
- Vacha, R., Buch, V., Milet, A., Devlin, J. P. & Jungwirth, P. Response to Comment on Autoionization at the surface of neat water: is the top layer pH neutral, basic, or acidic? by J. K. Beattie. *Phys. Chem. Chem. Phys.*, 2007, **9**, DOI: 10.1039/b713702h. *Phys. Chem. Chem. Phys.* **10**, 332–333 (2008).
- Petersen, P. B. & Saykally, R. J. Evidence for an enhanced hydronium concentration at the liquid water surface. *J. Phys. Chem. B* **109**, 7976–7980 (2005).
- Robertson, W. H., Diken, E. G., Price, E. A., Shin, J. W. & Johnson, M. A. Spectroscopic determination of the OH[−] solvation shell in the OH[−](H₂O)_n clusters. *Science* **299**, 1367–1372 (2003).
- Cappa, C. D. *et al.* Effects of alkali metal halide salts on the hydrogen bond network of liquid water. *J. Phys. Chem. B* **109**, 7046–7052 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the BESSY staff for assistance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.W. (bernd.winter@bessy.de).

LETTERS

The increasing intensity of the strongest tropical cyclones

James B. Elsner¹, James P. Kossin² & Thomas H. Jagger¹

Atlantic tropical cyclones are getting stronger on average, with a 30-year trend that has been related to an increase in ocean temperatures over the Atlantic Ocean and elsewhere^{1–4}. Over the rest of the tropics, however, possible trends in tropical cyclone intensity are less obvious, owing to the unreliability and incompleteness of the observational record and to a restricted focus, in previous trend analyses, on changes in average intensity. Here we overcome these two limitations by examining trends in the upper quantiles of per-cyclone maximum wind speeds (that is, the maximum intensities that cyclones achieve during their lifetimes), estimated from homogeneous data derived from an archive of satellite records. We find significant upward trends for wind speed quantiles above the 70th percentile, with trends as high as $0.3 \pm 0.09 \text{ m s}^{-1} \text{ yr}^{-1}$ (s.e.) for the strongest cyclones. We note separate upward trends in the estimated lifetime-maximum wind speeds of the very strongest tropical cyclones (99th percentile) over each ocean basin, with the largest increase at this quantile occurring over the North Atlantic, although not all basins show statistically significant increases. Our results are qualitatively consistent with the hypothesis that as the seas warm, the ocean has more energy to convert to tropical cyclone wind.

An important concern about the consequences of climate change is the potential increase in tropical cyclone activity. Theoretical arguments^{5,6} and modelling studies^{7,8} indicate that tropical cyclone winds should increase with increasing ocean temperature. Direct observational verification of this relationship over the global tropics is lacking, but Atlantic sea surface temperature (SST), which is correlated with global mean near-surface air temperature, helps explain¹ the recent upswing in frequency and intensity of Atlantic tropical cyclones. However, it has been argued that the data are not reliable enough to make assertions about the relationship between climate change and hurricanes^{9–13} and that the correlation may involve both regional and remote SSTs^{14,15}. Here we shed new light on this topic by using globally consistent satellite-derived tropical cyclone wind speeds¹⁶ and by focusing on the lifetime-maximum wind speeds of the strongest tropical cyclones each year.

Figure 1a shows the satellite-derived lifetime-maximum wind speeds grouped by year over the period 1981–2006, displayed as box plots (see Supplementary Information). The number of cyclones per year over the globe is shown above the time axis; there is no trend in these counts. Also, there is no trend in the median lifetime-maximum wind speed, as shown by the nearly horizontal red line, which is the best-fit line through the annual 50th-percentile values (black dashes inside the boxes). However at cyclone wind speeds above the median, upward trends are noted. Thus, the upper-quartile value (top of the box) is increasing (green line) and so are higher quantile values (for example the top of the vertical dashed line), where the upward trends are more pronounced.

To quantify and determine the significance of these trends, we use quantile regression. Quantile regression as employed here is a method to estimate the change (trend) in lifetime-maximum wind speed quantile as a function of year. A quantile is a point taken from

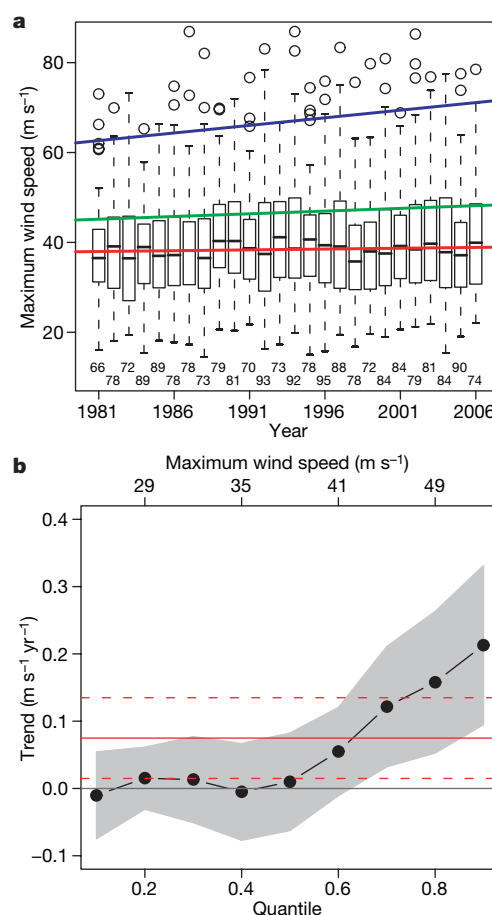


Figure 1 | Analysis and model results of satellite-derived tropical cyclone lifetime-maximum wind speeds. **a**, Box plots by year. Trend lines are shown for the median, 0.75 quantile, and 1.5 times the interquartile range. **b**, Trends in global satellite-derived tropical cyclone maximum wind speeds by quantile, from 0.1 to 0.9 in increments of 0.1. Trends are estimated coefficients from quantile regression in units of metres per second per year. The point-wise 90% confidence band is shown in grey, under the assumption that the errors are independent and identically distributed. The solid red line is the trend from a least-squares regression of wind speed as a function of year and the dashed red lines delineate the 90% point-wise confidence band about this trend.

¹Department of Geography, Florida State University, Tallahassee, Florida 32306, USA. ²Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.

the inverse cumulative distribution function of the set of wind speeds so that, for example, the 0.7 quantile is the value such that 70% of the tropical cyclones have lifetime-maximum wind speeds below this value (70th percentile).

Figure 1b shows global trends in tropical cyclone lifetime-maximum wind speeds for selected quantiles. Trends are near zero for the lower quantiles (median and below), but are upward for the higher quantiles, with the largest trends noted for the highest quantile (90th percentile). The shading shows the 90% point-wise confidence band about these trend estimates. Trends significantly above zero are seen for quantiles above 0.7. The maximum wind speeds over the entire period of record corresponding to the selected quantiles are also displayed. For comparison, the red lines are from a least-squares regression of maximum wind speed as a function of year, with the solid line showing the trend of the mean lifetime-maximum wind speed and the dashed lines indicating the 90% confidence limits about this trend. We note that the trend value of approximately $0.15 \text{ m s}^{-1} \text{ yr}^{-1}$ interpolated from Fig. 1b at the 75th percentile matches the slope value of the trend line corresponding to the upper quantiles shown as the green line in Fig. 1a. The results clearly show that the strongest tropical cyclones are getting stronger.

To examine whether these global increases are the result of trends occurring in one or two tropical cyclone basins, we use quantile regression to model the satellite-derived wind speeds from each basin separately (Fig. 2). With the exception of the South Pacific Ocean, all tropical cyclone basins show increases in the lifetime-maximum wind speeds of the strongest storms. The increases are greatest for cyclones over the North Atlantic and northern Indian oceans. Consistent with earlier results, only over the North Atlantic is there a significant increase in average tropical cyclone intensity. The width of the confidence interval is inversely proportional to the number of cyclones and proportional to the variability in wind speeds. We

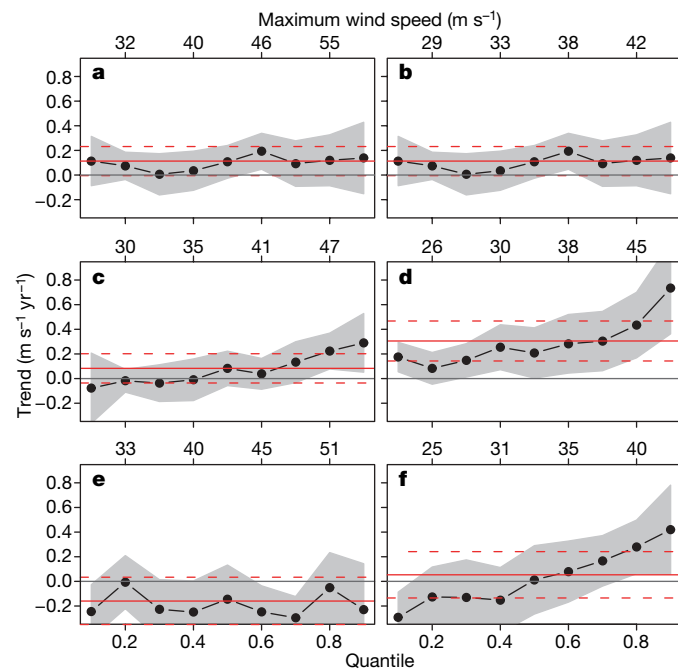


Figure 2 | Trends in satellite-derived tropical cyclone lifetime-maximum wind speeds from quantile regression. **a**, Western North Pacific Ocean cyclones (cyclone count, 698); **b**, eastern North Pacific Ocean cyclones (423); **c**, southern Indian Ocean cyclones (413); **d**, North Atlantic Ocean cyclones (291); **e**, South Pacific Ocean cyclones (157); **f**, northern Indian Ocean cyclones (115). The point-wise 90% confidence band is shown in grey, under the assumption that the errors are independent and identically distributed. Each solid red line is the trend from a least-squares regression of wind speed as a function of year and the dashed red lines delineate the 90% point-wise confidence band about this trend.

display trends and associated standard errors and P values for upper-quantile (≥ 85 th-percentile) lifetime-maximum wind speeds in Table 1. We note significant ($P < 0.05$) increases for at least one quantile level in all six basins, and upward trends in the wind speeds of the strongest tropical cyclones in all basins for the highest quantile considered (99th percentile), although not all trends at this extreme quantile are statistically significant.

The potential intensity of a tropical cyclone is directly related to SST below the cyclone, all else being equal^{5,6,17,18}. Because the strongest cyclones at their maxima are, on average, closest to their maximum potential intensities, increases in observed maximum wind speeds should occur with SST at the upper quantiles. To test this, we averaged Hadley Centre¹⁹ SST data over each of the six tropical cyclone basins during the peak months of their respective tropical cyclone seasons. The basin means are then averaged to obtain a global tropics SST value for each year over the period 1981–2006. These values are subsequently used instead of year in the quantile regression. The results are shown in Fig. 3a. Consistent with the theory, the trends in units of metres per second per degree Celsius are positive for the upper quantiles. For a 1°C rise in SST, the results show an increase of $1.9 \pm 2.9 \text{ m s}^{-1}$ (s.e.) in the value of the 80th percentile and $6.5 \pm 4.2 \text{ m s}^{-1}$ in the value of the 90th percentile.

Table 1 | Summary statistics

Statistic	Quantile				
	0.85	0.90	0.95	0.975	0.99
Global (2,097)					
$W (\text{m s}^{-1})$	51.9	55.8	62.6	68.8	75.9
Trend ($\text{m s}^{-1} \text{ yr}^{-1}$)	+0.19	+0.21	+0.18	+0.25	+0.30
s.e. ($\text{m s}^{-1} \text{ yr}^{-1}$)	0.049	0.072	0.141	0.122	0.093
P	<0.001	0.003	0.212	0.044	0.001
Western North Pacific (698)					
$W (\text{m s}^{-1})$	58.9	63.7	69.6	73.1	77.6
Trend ($\text{m s}^{-1} \text{ yr}^{-1}$)	+0.12	+0.14	+0.09	+0.29	+0.36
s.e. ($\text{m s}^{-1} \text{ yr}^{-1}$)	0.190	0.177	0.192	0.116	0.230
P	0.520	0.434	0.647	0.012	0.115
Eastern North Pacific (423)					
$W (\text{m s}^{-1})$	44.8	48.8	53.0	57.3	62.8
Trend ($\text{m s}^{-1} \text{ yr}^{-1}$)	+0.11	+0.16	+0.33	+0.46	+0.80
s.e. ($\text{m s}^{-1} \text{ yr}^{-1}$)	0.117	0.156	0.160	0.184	NA
P	0.327	0.308	0.042	0.014	NA
Southern Indian (413)					
$W (\text{m s}^{-1})$	49.2	51.3	56.4	57.8	62.6
Trend ($\text{m s}^{-1} \text{ yr}^{-1}$)	+0.28	+0.29	+0.44	+0.43	+0.69
s.e. ($\text{m s}^{-1} \text{ yr}^{-1}$)	0.065	0.145	0.177	0.178	0.353
P	<0.001	0.046	0.014	0.016	0.052
North Atlantic (291)					
$W (\text{m s}^{-1})$	48.9	54.8	60.3	72.7	77.8
Trend ($\text{m s}^{-1} \text{ yr}^{-1}$)	+0.63	+0.73	+0.81	+1.11	+1.52
s.e. ($\text{m s}^{-1} \text{ yr}^{-1}$)	0.228	0.226	0.449	0.356	NA
P	0.006	0.001	0.073	0.002	NA
South Pacific (157)					
$W (\text{m s}^{-1})$	52.8	54.7	59.8	65.8	67.1
Trend ($\text{m s}^{-1} \text{ yr}^{-1}$)	−0.04	−0.23	−0.37	−0.06	+0.46
s.e. ($\text{m s}^{-1} \text{ yr}^{-1}$)	0.174	0.227	0.214	0.563	0.210
P	0.803	0.313	0.088	0.914	0.030
Northern Indian (115)					
$W (\text{m s}^{-1})$	41.5	45.0	47.4	50.4	56.4
Trend ($\text{m s}^{-1} \text{ yr}^{-1}$)	+0.48	+0.42	+0.69	+0.83	+0.87
s.e. ($\text{m s}^{-1} \text{ yr}^{-1}$)	0.224	0.220	0.222	NA	NA
P	0.034	0.059	0.002	NA	NA

Statistics are from a quantile regression of lifetime-maximum tropical cyclone wind speed (derived from satellites) as a function of year, either globally or by tropical cyclone basin. Sample size (number of tropical cyclones) is given in parentheses next to the basin name. Values are shown for selected upper quantiles (0.85, 0.90, 0.95, 0.975, and 0.99). For each quantile, W denotes the tropical cyclone lifetime-maximum wind speed over all cyclones in the basin and over all years in the analysis (1981–2006). For some extreme quantiles the s.e., computed assuming independent and identically distributed errors, and P value are not reliable and so are reported as not available (NA).

Upward trends can be interpreted as an increase in the number of cyclones exceeding a threshold quantile. For example, at the 80th percentile, on average 17 cyclones globally exceed 49 m s^{-1} . With a 1°C rise in SST, the 80th percentile increases to 51 m s^{-1} . At this threshold level, on average 13 cyclones per year are observed. So the increase in SST of 1°C results in an increase in the global frequency of strong cyclones from 13 to 17 cyclones (31%) per year. The best estimates indicate that the strongest tropical cyclones are getting stronger with increasing SST, but the uncertainty ranges are relatively large. The relationship does not imply causality and is not directly comparable to results from numerical models with forced SST changes. We make no attempt here to control other factors probably related to intense tropical storminess such as changes in region of origin, cyclone duration, El Niño conditions and solar activity.

For comparison, we repeat the quantile regression for the set of lifetime-maximum wind speeds based on the global best-track data sets (Fig. 3b). The best-track data represent a best estimate of cyclone position and intensity from all available information. Results are similar, showing an increase in lifetime-maximum wind speed per degree Celsius for the set of strongest cyclones but not for the set of weaker cyclones. Magnitudes of the change are not directly comparable, because the variance of the satellite-derived wind speeds is, by construction (regression model), less than observed wind speeds, which results in a mismatch in the quantile values. Moreover, inhomogeneities in the best-track data due to changes in the availability

and quality of information over time probably contribute to the magnitude of this trend.

Recent results from the analyses of global tropical cyclone trends have been questioned owing to a lack of consensus regarding the reliability of the data. Moreover, results have not been matched to theory, because the focus was on a change in mean tropical cyclone statistics. In contrast, the results presented here are conclusive in showing significant increasing trends in the satellite-derived lifetime-maximum wind speeds of the strongest tropical cyclones globally, and are qualitatively consistent with the heat-engine theory of cyclone intensity. Thus, as seas warm, the ocean has more energy that can be converted to tropical cyclone wind.

Regional differences in the magnitude of the upward trends are possibly due in part to the rate of warming relative to the existing warmth in the basin. Relatively cooler basins with large SST increases should therefore show the greatest upward trends in the intensity of the strongest tropical cyclones. The three coolest basins over the period 1981–2006 are the North Atlantic (27.6°C), the eastern North Pacific (27.9°C) and the southern Indian (27.5°C), and the rates of warming in these basins are respectively $0.69 \pm 0.18^\circ\text{C}$, $0.33 \pm 0.24^\circ\text{C}$ and $0.21 \pm 0.16^\circ\text{C}$ per 30 years. These basins show corresponding upward trends at the 99th percentile of 1.52, 0.80 and $0.69 \text{ m s}^{-1}^\circ\text{C}^{-1}$, respectively. Small positive correlations are noted between the warming trends in the tropical oceans and the upward trends in the intensity of the strongest tropical cyclones, using all six basins, with the largest correlation ($r = 0.47$, $N = 6$) occurring for the 99th-percentile trends. It is necessary to control other factors such as changes in upper-tropospheric temperatures, shearing winds and proximity to land to better understand regional differences in these trends.

METHODS SUMMARY

We use log-linear regression to model the lifetime maximum wind speeds using principal components of brightness temperature profiles from satellite imagery^{20–24} for 171 tropical cyclones over the North Atlantic Ocean. The regression model is modified from ref. 16 to better account for the skewness in wind speed values. Model details and diagnostics are given in the Supplementary Information. We apply the regression model to satellite imagery for 2,097 tropical cyclones around the globe over the period 1981–2006 to produce the satellite-derived per-cyclone lifetime-maximum wind speeds.

We subsequently estimate trends in satellite-derived lifetime-maximum wind speeds using quantile regression. Quantile regression extends ordinary least-squares regression to conditional quantiles of the response variable²⁵ (lifetime-maximum wind speed). Quantiles are values taken at regular intervals from the cumulative distribution function. The quantiles divide a set of ordered wind speeds into equally sized subsets. A minimization procedure determines the quantile regression trend. More details on quantile regression are given in the Supplementary Information. All statistics are performed using the software environment R (<http://www.r-project.org>) and the quantile regression package *quantreg* (R package version 4.17; <http://www.r-project.org>).

Received 25 January; accepted 27 June 2008.

1. Emanuel, K. A. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **436**, 686–688 (2005).
2. Webster, P. J., Holland, G. J., Curry, J. A. & Chang, H. R. Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science* **309**, 1844–1846 (2005).
3. Trenberth, K. Uncertainty in hurricanes and global warming. *Science* **308**, 1753–1754 (2005).
4. Elsner, J. B. Granger causality and Atlantic hurricanes. *Tellus* **59A**, 476–485 (2007).
5. Emanuel, K. A. The theory of hurricanes. *Annu. Rev. Fluid Mech.* **23**, 179–196 (1991).
6. Holland, G. J. The maximum potential intensity of tropical cyclones. *J. Atmos. Sci.* **54**, 2519–2541 (1997).
7. Knutson, T. R. & Tuleya, R. E. Impact of CO_2 -induced warming on simulated hurricane intensity and precipitation: Sensitivity to the choice of climate model and convective parameterization. *J. Clim.* **17**, 3477–3495 (2004).
8. Bengtsson, L. *et al.* How may tropical cyclones change in a warmer climate. *Tellus* **59A**, 539–561 (2007).
9. Pielke, R. A. Jr, Landsea, C., Mayeld, M., Laver, J. & Pasch, R. Hurricanes and global warming. *Bull. Am. Meteorol. Soc.* **86**, 1571–1575 (2005).

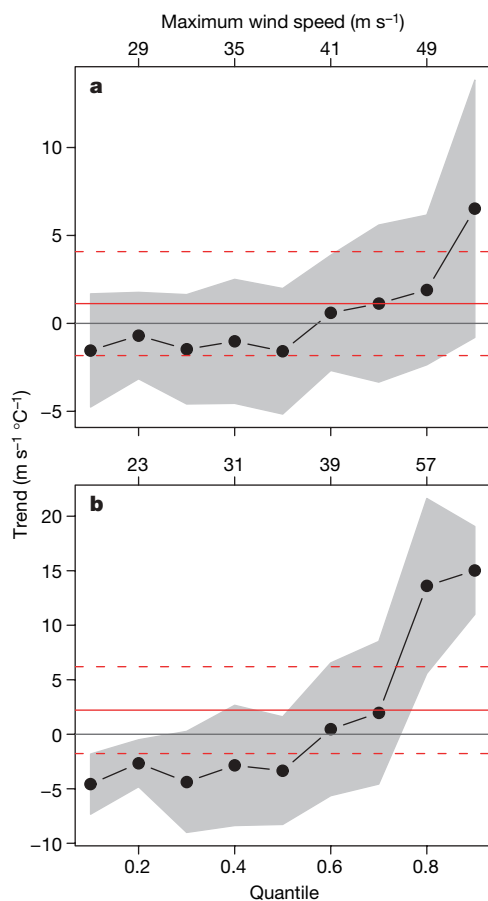


Figure 3 | Quantile regression of tropical cyclone lifetime-maximum wind speed on globally averaged tropical storm basin SST. **a**, Plotted using satellite-derived maximum wind speeds; **b**, plotted using maximum wind speeds as recorded in the best-track observational archives. The point-wise 90% confidence band is shown in grey under the assumption that the errors are independent and identically distributed. Each solid red line is the trend from a least-squares regression of wind speed as a function of year and the dashed red lines delineate the 90% point-wise confidence band about this trend.

10. Landsea, C. W. Hurricanes and global warming. *Nature* **438**, E11–E13 (2005).
11. Chan, J. C. L. Comment on “Changes in tropical cyclone number, duration, and intensity in a warming environment”. *Science* **311**, 1713 (2006).
12. Klotzbach, P. J. Trends in global tropical cyclone activity over the past twenty years (1986–2005). *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL025881 (2006).
13. Landsea, C. W., Harper, B. A., Hoarau, K. & Knaff, J. A. Can we detect trends in extreme tropical cyclones? *Science* **313**, 452–454 (2006).
14. Vecchi, G. A. & Soden, B. J. Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature* **450**, 1066–1070 (2007).
15. Swanson, K. L. Nonlocality of Atlantic tropical cyclone intensities. *Geochim. Geophys. Geosyst.* **9**, doi:10.1029/2007GC001844 (2008).
16. Kossin, J. P., Knapp, K. R., Vimont, D. J., Murnane, R. J. & Harper, B. A. A globally consistent reanalysis of hurricane variability and trends. *Geophys. Res. Lett.* **34**, doi:10.1029/2006GL028836 (2007).
17. Shen, W., Tuleya, R. E. & Ginis, I. A sensitivity study of the thermodynamic environment on GFDL model hurricane intensity: Implications for global warming. *J. Clim.* **13**, 109–121 (2000).
18. Bister, M. & Emanuel, K. A. Low frequency variability of tropical cyclone potential intensity. 1. Interannual to interdecadal variability. *J. Geophys. Res.* **107**, doi:10.1029/2001JD000776 (2002).
19. Rayner, N. A. *et al.* Global analysis of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, doi:10.1029/2002JD002670 (2003).
20. Knapp, K. R. Calibration assessment of ISCCP geostationary infrared observations using HIRS. *J. Atmos. Oceanic Technol.* **25**, 183–195 (2008).
21. Knapp, K. R. & Kossin, J. P. A new global tropical cyclone data set from ISCCP B1 geostationary satellite observations. *J. Appl. Remote Sensing* **1**, 013505 (2007).
22. Knapp, K. R., Bates, J. J. & Barkstrom, B. Scientific data stewardship: Lessons learned from a satellite-data rescue effort. *Bull. Am. Meteorol. Soc.* **88**, 1359–1361 (2007).
23. Franklin, J. L., Black, M. L. & Valde, K. GPS dropwindsonde wind profiles in hurricanes and their operational implications. *Weather Forecast.* **18**, 32–44 (2003).
24. Gunshor, M. M., Schmit, T. J. & Menzel, W. P. Intercalibration of the infrared window and water vapor channels on operational geostationary environmental satellites using a single polar-orbiting satellite. *J. Atmos. Oceanic Technol.* **21**, 61–68 (2004).
25. Yu, K., Lu, Z. & Stander, J. Quantile regression: applications and current research areas. *Statistician* **52**, 331–350 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The work was supported by the US National Science Foundation (ATM-0738172 and ATM-0614812) and by the Risk Prediction Initiative of the Bermuda Institute for Ocean Studies (RPI06-3-001).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.B.E. (jelsner@fsu.edu).

LETTERS

Multimodal warning signals for a multiple predator world

John M. Ratcliffe^{1*} & Marie L. Nydam^{2*}

Aposematism is an anti-predator defence, dependent on a predator's ability to associate unprofitable prey with a prey-borne signal¹. Multimodal signals should vary in efficacy according to the sensory systems of different predators; however, until now, the impact of multiple predator classes on the evolution of these signals had not been investigated^{2,3}. Here, using a community-level molecular phylogeny to generate phylogenetically independent contrasts, we show that warning signals of tiger moths vary according to the seasonal and daily activity patterns of birds and bats—predators with divergent sensory capacities. Many tiger moths advertise chemical defence^{4,5} using conspicuous colouration and/or ultrasonic clicks^{3,6}. During spring, when birds are active and bats less so, we found that tiger moths did not produce ultrasonic clicks. Throughout both spring and summer, tiger moths most active during the day were visually conspicuous. Those species emerging later in the season produced ultrasonic clicks; those that were most nocturnal were visually cryptic. Our results indicate that selective pressures from multiple predator classes have distinct roles in the evolution of multimodal warning displays now effective against a single predator class. We also suggest that the evolution of acoustic warning signals may lack the theoretical difficulties associated with the origination of conspicuous colouration.

Insectivorous birds and bats are major predators of adult Lepidoptera. In south-eastern Ontario, Canada, where the field data for this study were collected, residential and migratory insect-eating birds actively forage before, during and after the seasonal emergence of tiger moths^{7–9}; a slight plateau in bird abundance occurs between early June and early July^{7,8}. Conversely, peak bat foraging activity does not occur until early July and lasts until mid August. In early May, bat foraging activity is at ~15% of its peak, rising to only ~50% by late June^{10–12}. Most insectivorous birds are diurnal predators sensitive to wavelengths extending beyond the human visual spectrum to the ultraviolet^{13,14}. Vespertilionid bats are nocturnal, with their scotopic vision and poor visual acuity unsuited to the discrimination of insect prey¹⁵. Instead, these bats detect and locate prey and other objects in their immediate environment using the echoes returning from their mostly ultrasonic calls (>20 kHz). Birds do not echolocate prey nor are their ears sensitive to frequencies above 10 kHz (their range of best frequency is 2–5 kHz)¹⁶. However, both insect-eating birds and bats readily learn taste aversions to novel prey when prey cues are associated with toxicity^{17,18}. This adaptive specialization of learning may be a necessary precondition in predators for the evolution of aposematic signalling in prey¹⁷.

Many tiger moth species are unpalatable to birds and bats^{3–6,15}. All species possess bat-detecting ears, and some respond to aerial hawking bats with ultrasonic clicks from sound-producing organs known as tymbals^{6,10,15,19}. Many are also visually conspicuous^{3,4,15,20}, which

may allow tiger moths to be more diurnal than most moths²⁰. For each of the 26 species included in our study, colouration/pattern were scored as low contrast (cryptic; v1), white (conspicuous; v2) or high contrast (conspicuous; v3) (Fig. 1). Each species was also scored as

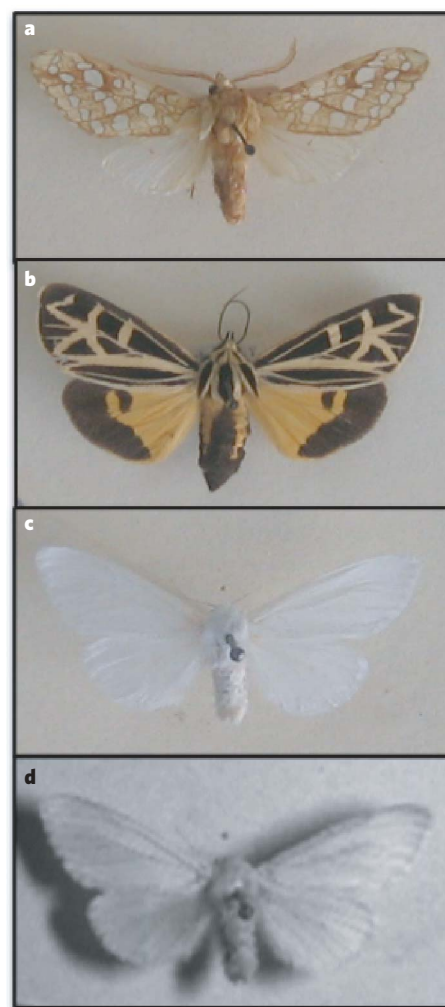


Figure 1 | Representatives of each visual class. **a**, Low colour contrast: *Lophocampa caryae*. **b**, High colour contrast: *Grammia anna*. **c**, White: *Hyphantria cunea*. **d**, Ultraviolet photograph of *H. cunea*, one of the 37 moth species scored as not exhibiting a qualitative difference in pattern between colour (**c**) and ultraviolet (**d**) photographs. Only *Lycomorpha pholus* was noted as having a different pattern under ultraviolet light (see Methods).

¹Center for Sound Communication, Institute of Biology, University of Southern Denmark, DK-5230 Odense M, Denmark. ²Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA.

*These authors contributed equally to this work.

being silent (a1), a simple-ultrasound producer (a2) or a complex-ultrasound producer (a3) (Fig. 2). The emergence date for each species was taken as the date of first capture at or in the vicinity of the study site⁹. Diel flight periodicity (DFP, 24-h activity pattern) for each species was determined using a previously reported behavioural assay²⁰ and percentage nocturnality was calculated using a site-specific rubric²⁰ (for details of trait quantification, see Methods Summary and Methods).

We built a community-level molecular phylogeny using one mitochondrial (*cytochrome oxidase I*, *COI*) locus and two nuclear (elongation factor 1a, *EF1a*, and *wingless*) loci (Fig. 3, see Methods Summary and Methods). For each species, character values for each of the four traits (emergence date, percentage nocturnality, visual class, acoustic class) were mapped onto this phylogeny (Fig. 3), which we then used to perform phylogenetically independent contrasts using Comparative Analysis of Independent Contrasts (CAIC)²¹ version 2.6.9 (see Methods). Percentage nocturnality (hereafter, nocturnality) was not related to the emergence date ($F_{1,24} = 1.39$, $r^2 = 0.06$, $P = 0.25$). The emergence date (hereafter, emergence) was not related to the visual category (Fig. 4a, analysis of variance, ANOVA $F_{2,22} = 1.601$, $P = 0.209$). Nocturnality was significantly higher in low-contrast species than in white and high-contrast species, which did not differ significantly from one another (Fig. 4b, ANOVA $F_{2,22} = 9.546$, $P < 0.001$; Tukey HSD (honestly significant difference) post-hoc comparisons: low contrast versus white, $P = 0.001$; low contrast versus high contrast, $P = 0.001$; white versus high contrast, $P = 0.985$). Emergence occurred significantly later in the season for sound-producing species than for silent species (Fig. 4c, ANOVA $F_{2,22} = 5.593$, $P = 0.006$; Tukey HSD post-hoc comparisons: silent versus complex, $P = 0.005$; silent versus simple, $P = 0.059$; simple versus complex, $P = 0.63$). Nocturnality was not related to the acoustic category (Fig. 4d, ANOVA $F_{2,22} = 0.534$, $P = 0.588$). We used the program Mesquite version 2.0 (ref. 22) to investigate potential phylogenetic correlations between sound production and colouration considered as binary traits (silent (a1) or sound-producing (a2 + a3); low contrast (cryptic; v1) or white/high contrast (conspicuous; v2 + v3)) and found none ($P = 0.726$).

Differences in insectivorous bird and bat daily and seasonal activity patterns in the Nearctic allow for the use of DFP and emergence as proxies for species-specific differences in moths' exposure to these

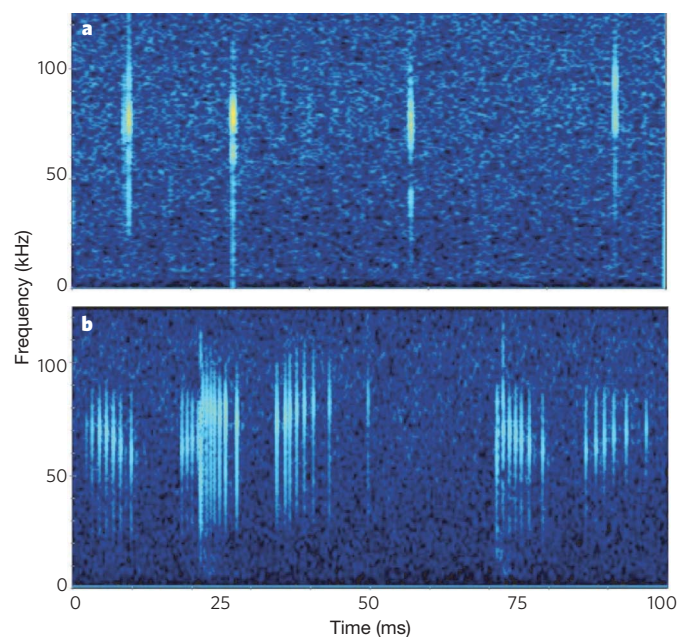


Figure 2 | Representative sonograms. **a**, Simple-ultrasound producers (for example, *Cisseps fulvicollis*). **b**, Complex-sound producers (for example, *Cynia tenera*).

two predator classes. The correlation between reduced nocturnality and both classes of conspicuous colouration (Fig. 4b), and the lack of correlation between the latter and emergence (Fig. 4a), is as predicted by the relatively stable seasonal abundance of diurnal, vision-dependent insect-eating birds. Conversely, the positive correlation between emergence and sound production (Fig. 4c) is as predicted by the increasing selective pressure put on moths by nocturnal, echolocating bats as the season progresses¹⁰. The lack of correlation between nocturnality and sound production may be because, except for *Lycomorpha pholus*, all species are greater than 40% nocturnal and 19 out of 26 species are 60% or more (Fig. 3). The significant difference in emergence between silent and complex-sound-producing tiger moths corroborates evidence that, for bats, complex sounds are more salient warning signals than those directed at other sensory modalities^{5,15,23}. More than this, the lack of positive correlation between sound production and conspicuous colouration argues against the evolution of multimodal warning signals for the function of improving learning and memory in either single predator class.

Warning signals function by informing would-be predators that the sender is unprofitable as prey¹⁻³. The visual and acoustic warning signals of tiger moths have been shown to be readily associated with toxicity in birds and bats, respectively^{3,5,23}. However, most visual aposematic signals are continuously displayed²; in short, they are easily seen. The fixation of an initially rare visually conspicuous phenotype is therefore paradoxical^{2,3}. One possible solution is that predator avoidance of novel foods and reluctance to add these foods to the diet—behaviours common among birds³—may allow initially rare phenotypes to become more frequent in the population^{2,3}. But because many bats are cavalier with respect to diet selection, attacking muted tiger

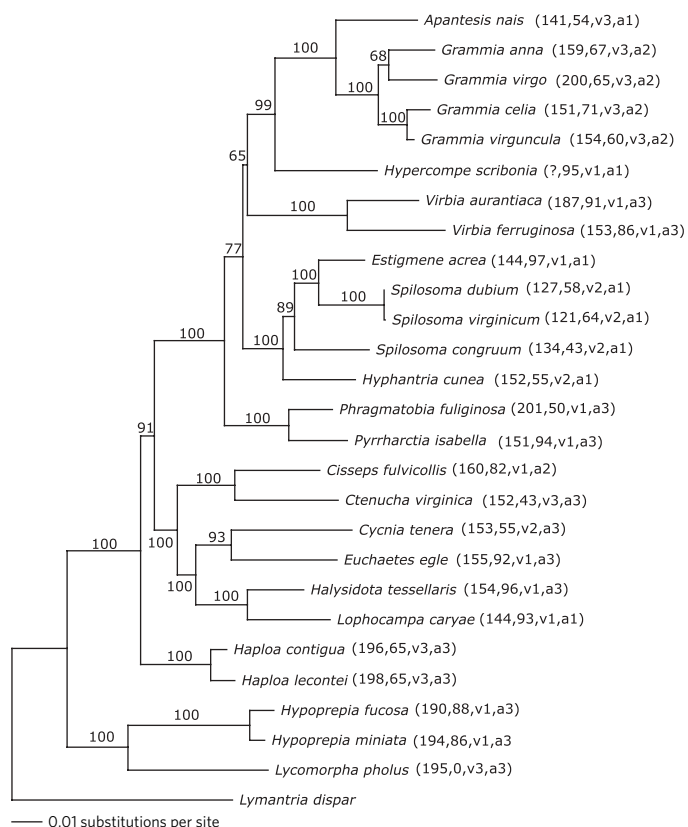


Figure 3 | 50% majority consensus phylogram of the bayesian trees. Numbers along the branches are bayesian posterior probabilities. Within parentheses after each species name are the character traits for that species: emergence (for example, 30 June 2007 = 181st day of the year), percentage nocturnality, visual class and acoustic class. All combinations of visual and acoustic signal classes were observed except for white, simple-sound-producing moths (v2,a2).

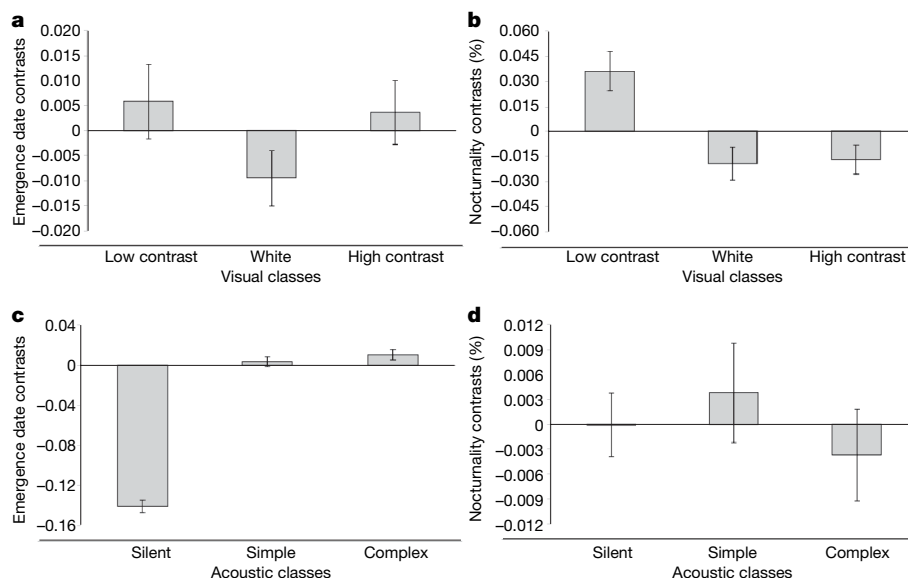


Figure 4 | Phylogenetically independent contrasts (mean \pm s.e.).

a, Emergence date contrasts (calculated in CAIC v. 2.6.9) in relation to three visual classes (low colour contrast, white, high colour contrast).

b, Percentage nocturnality contrasts in relation to three visual classes.

moths and moth-sized inanimate objects under otherwise natural conditions^{24,25}, the ability of rare phenotypes to increase due to predators' wariness^{2,3} may not apply during bat–moth interaction. However, unlike visual signals, during these interactions the ultrasonic clicks of tiger moths are elicited only by the echolocation calls of, or contact with, an attacking bat^{15,19,26}. These signals are thus 'invisible' to bats at a distance and are produced by a tiger moth only after a bat has already detected it and is on course for interception; they therefore may be exempt from the rare/conspicuous paradox.

Our study is one of few to investigate the evolution of aposematic signalling using a phylogenetic framework^{2,3}, and is the first, to our knowledge, to consider the evolution of multimodal displays selected by multiple predators. Multiple sensory signals have been suggested to act synergistically to reinforce aversion learning^{3,27}, and the visual and acoustic aposematic signals of tiger moths and other protected prey have been suggested to have evolved to serve a single, synergistic function²⁷. Indeed, sounds audible to chicks, *Gallus gallus domesticus*, can improve visual discrimination learning^{3,28}, and the lower periphery of the frequencies found in some tiger moth clicks may be audible to birds at very close range (that is, when held in the bird's beak). However, in our system only two sound-producing moths are >50% diurnal (Fig. 3). At least in the case of tiger moths, this dearth of diurnal sound producers provides evidence against multiple signals having initially evolved in response to selective pressures from a single predator class and/or that, once evolved, they are maintained by selective pressures from a single predator class. Taken together, our results suggest that the proximate benefits of some multimodal displays are not reflective of their evolutionary histories. These histories may be better understood in the context of selective pressures from multiple predator classes—classes defined by their own sensory capacities and life history traits.

METHODS SUMMARY

All bat, bird and moth activity and emergence data were collected at or near Queen's University Biological Station in south-eastern Ontario, Canada^{6–12,20}. For DFP, we used previously reported data²⁰ for 10 out of the 26 tiger moth species, and used the same setup and design to collect data for the remaining 16 species ($N = 4$ per species). For acoustic classification, data for all but three species were taken from the literature^{6,10,15,19,23,24,26}. Species that produced sounds at a maximum rate of <100 clicks per s were classified as 'simple', and those that produced sounds at a maximum rate of >500 clicks per s were classified as

c, Emergence date contrasts in relation to three acoustic classes (silent, simple-sound producers, complex-sound producers). **d**, Percentage nocturnality contrasts in relation to three acoustic classes.

'complex'. No species had a maximum rate between 100 and 500 sounds per s. *Hypercompe scribonia* did not produce sounds and was scored as 'silent'. Sounds produced by *Ciseps fulvicollis* and *Ctenucha virginica* were recorded and analysed as described elsewhere¹⁵. For visual classification, digital colour and ultraviolet photographs (custom setup described elsewhere²⁹) of spread specimens of the 26 tiger moth species and 12 other sympatric noctuid species were analysed and classified using a computer-driven routine (see Methods). We asked human subjects ($N = 15$) to compare colour images to ultraviolet images to determine whether qualitative pattern changes exist between these sets of images (Fig. 1c, d). For phylogenetic inference, portions of one mitochondrial (*COI*) and two nuclear (*EF1a* and *wingless*) genes totalling ~2 kilobases were amplified and sequenced (see Methods). The topology and branch lengths of a 50% majority consensus phylogenetic tree constructed using MrBayes 3.1.2 (ref. 30) were used to calculate standardized linear contrasts using the program CAIC²¹. All comparative analyses reported herein use actual branch lengths.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 February; accepted 14 May 2008.

- Cott, H. B. *Adaptive Coloration in Animals* (Methuen, London, 1940).
- Mappes, J., Marples, N. & Endler, J. A. The complex business of survival by aposematism. *Trends Ecol. Evol.* **20**, 598–603 (2005).
- Ruxton, G. D., Sherratt, T. N. & Speed, M. *Avoiding Attack: the Evolutionary Ecology of Crypsis, Warning Signals and Mimicry* (Oxford Univ. Press, Oxford, 2004).
- Nishida, R. Sequestration of defensive substances from plants by Lepidoptera. *Annu. Rev. Entomol.* **47**, 57–92 (2002).
- Hristov, N. & Conner, W. E. Effectiveness of tiger moth (Lepidoptera, Arctiidae) chemical defenses against an insectivorous bat (*Eptesicus fuscus*). *Chemoecology* **15**, 105–113 (2005).
- Hristov, I. H. & Conner, W. E. Sound strategy: acoustic aposematism in the bat–tiger moth arms race. *Naturwissenschaften* **92**, 164–169 (2005).
- Cadman, M. D., Eagles, P. F. J. & Helleiner, F. M. *Atlas of the Breeding Birds of Ontario* (Univ. Waterloo Press, Waterloo, 1987).
- Godfrey, W. E. 1986. *The Birds of Canada*, rev. ed. (Nat'l Mus. Nat. Sci. Canada, Ottawa, 1986).
- Ward, P. S., Harmsen, R. & Hebert, P. D. N. An annotated checklist of the Macrolethocera of South-eastern Ontario. *J. Res. Lepidoptera* **13**, 23–42 (1974).
- Fullard, J. H. Phenology of sound-producing arctiid moths and the activity of insectivorous bats. *Nature* **267**, 42–43 (1977).
- Fullard, J. H. & Barclay, R. M. R. Audition in spring moths as a possible response to differential levels of insectivorous bat predation. *Can. J. Zool.* **58**, 1745–1750 (1980).
- Yack, J. E. Seasonal partitioning of atympanate moths in relation to bat activity. *Can. J. Zool.* **66**, 753–755 (1988).
- Bennett, A. T. D. & Cuthill, I. C. Ultraviolet vision in birds: what is its function? *Vision Res.* **34**, 1471–1478 (1994).

14. Lyttinen, A., Lindstrom, L. & Mappes, J. Ultraviolet reflection and predation risk in diurnal and nocturnal Lepidoptera. *Behav. Ecol.* **15**, 982–987 (2004).
15. Ratcliffe, J. M. & Fullard, J. H. The adaptive function of tiger moth clicks against echolocating bats: an experimental and synthetic approach. *J. Exp. Biol.* **208**, 4689–4698 (2005).
16. Dooling, R. J., Lohr, B. & Dent, M. L. in *Comparative Hearing: Birds and Reptiles* (eds Dooling, R. J., Popper, A. N. & Fay, R. R.) 308–359 (Springer, New York, 2000).
17. Nicolaus, L. K., Cassel, J. F., Carlson, R. B. & Gustavson, C. R. Taste-aversion conditioning of crows to control predation on eggs. *Science* **220**, 212–214 (1983).
18. Ratcliffe, J. M., Fenton, M. B. & Galef, B. G. An exception to the rule: common vampire bats do not learn taste aversions. *Anim. Behav.* **65**, 385–389 (2003).
19. Fullard, J. H. & Fenton, M. B. Acoustic and behavioural analyses of the sounds produced by some species of Nearctic Arctiidae (Lepidoptera). *Can. J. Zool.* **55**, 1213–1224 (1977).
20. Fullard, J. H. & Napoleone, N. Diel flight periodicity and the evolution of auditory defences in the Macrolepidoptera. *Anim. Behav.* **62**, 349–368 (2001).
21. Purvis, A. & Rambaut, A. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Comp. Appl. Biosci.* **11**, 247–251 (1995).
22. Maddison, W. P. & Maddison, D. R. *Mesquite: a modular system for evolutionary analysis* Version 2.0 <<http://mesquiteproject.org>> (2007).
23. Barber, J. R. & Conner, W. E. Acoustic mimicry in a predator–prey interaction. *Proc. Natl Acad. Sci. USA* **104**, 9331–9334 (2007).
24. Acharya, L. & Fenton, M. B. Echolocation behaviour of vespertilionid bats (*Lasiurus cinereus* and *L. borealis*) attacking airborne targets including arctiid moths. *Can. J. Zool.* **70**, 1292–1298 (1992).
25. Barclay, R. M. R. & Brigham, R. M. Constraints on optimal foraging: a field test of prey discrimination by echolocating insectivorous bats. *Anim. Behav.* **48**, 1013–1021 (1994).
26. Barber, J. R. & Conner, W. E. Tiger moth responses to a simulated bat attack: timing and duty cycle. *J. Exp. Biol.* **209**, 2637–2650 (2006).
27. Rowe, C. & Guilford, T. The evolution of multimodal warning displays. *Evol. Ecol.* **13**, 655–671 (1999).
28. Rowe, C. Sound improves visual discrimination learning in avian predators. *Proc. R. Soc. Lond. B* **269**, 1353–1357 (2002).
29. Gronquist, M. *et al.* Attractive and defensive functions of the ultraviolet pigments of a flower (*Hypericum calycinum*). *Proc. Natl Acad. Sci. USA* **98**, 13745–13750 (2001).
30. Ronquist, F. & Huelsenbeck, J. P. MrBayes3: Bayesian inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).

Acknowledgements We thank R. Hoebeke for access to the Cornell University Insect Collection, T. Eisner for the loan of ultraviolet photography equipment, and the 15 people who scored the photographs. J.M.R. thanks M. Fitzpatrick for discussion. J. Barber, C. Beatty, I. Cuthill, T. Eisner and J. Fullard provided detailed comments that improved the manuscript. This study was funded by research grants to J. Fullard (NSERC), R. Hoy (NIDCD) and J.M.R. (NSERC and DNSRC). Molecular work was conducted at Cornell in the laboratory of R. Harrison and the Evolutionary Genetics Core Facility.

Author Contributions M.L.N. was responsible for genetic analyses. J.M.R. was responsible for behavioural, comparative and signal analyses. J.M.R. wrote the manuscript.

Author Information All sequences were submitted to GenBank under accession numbers EU333575–EU333652. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.L.N. (mln32@cornell.edu) or J.M.R. (jmr@biology.sdu.dk).

METHODS

Colour quantification and visual categorization. Digital colour photographs of the 26 tiger moth species used in this study and of 12 sympatric noctuid species were taken outdoors under full sunlight using specimens from the Cornell University Insect Collection. We used a D40x 10.2 MP digital SLR camera with a 18–55 mm f/3.5–5.6G ED II AF-S DX Zoom-Nikkor Lens (Nikon Inc.). White balance was set manually using a Zebra 2-sided grey card (Novoflex). Shutter speed was adjusted for optimal exposure within a range of 1/125 s–1/500 s for each photograph while the aperture was set to F8 for the entire session. Each colour photograph was calibrated according to previously described methods³¹, using the six greys of the Gretag Macbeth Mini ColourChecker Chart (Gretag Macbeth AG/LLC), which was included in each moth photograph. All photographs were also calibrated globally to account for changes in sunlight during the photography session (that is, for colour photographs, Macbeth white was set equivalent to red (R) = green (G) = black (B) = 255; black was set equivalent to R = G = B = 0). TIFF files were exported to Image J v. 1.38x (National Institutes of Health) from which we saved *x,y* coordinates from the perimeter of each moth. DigitalColour Meter v. 3.4.1 (Apple Computer) was used to measure 8-bit (256-point) corrected RGB values for ten points for each species; sampled points were determined using the intersection of two randomly selected and matched vectors (each taken from randomly selected pairs of *x,y* coordinates). A single reflectance value was taken for each of the ten homologous points for each moth from the ultraviolet photographs. Each ultraviolet photograph was calibrated using Macbeth black (set equivalent to R = G = B = 0) and a Spectralon (Labsphere) white reflectance standard (set equivalent to R = G = B = 255); the chart and standard were included in each photograph. We input the four values (R, G, B, ultraviolet) for each of the ten points sampled for each species in the cluster analyses outlined below.

White moths were classified as such and were removed from further colour analysis for three reasons: first, white is conspicuous on many natural backgrounds; second, none exhibited pattern change under ultraviolet illumination (see Fig. 1c, d); and third, it is unclear whether white Lepidoptera possess warning colouration³. In JMP v. 7 (SAS Institute), we used hierarchical clustering (centroid, data not standardized) and derived distance values based on RGB/ultraviolet data for the remaining 33 species; using this same method, we input species distance values and found two clusters: cluster A (species of <170 distance units: all 'low contrast/cryptic' tiger and all 12 other noctuid species) and cluster B (species >200 distance units: all 'high contrast/conspicuous' tiger moth species). None of the moths in cluster B (high contrast/conspicuous) had patterns indicative of disruptive colouration³.

Phylogenetic inference and comparative analyses. Genomic DNA of one individual of each of the 26 ingroup species and of 1 outgroup species (*Lymantria dispar*) was extracted from the flight muscle tissue of whole moths stored in 95% ethanol using a Qiagen DNeasy Tissue Kit. 764 bp from mitochondrial *COI* and tRNA-leucine were amplified using the primers C1-J-2183 (alias Jerry), TL2-N-3014 (alias Pat), C1-J-2195 (alias CO1RLR) and TL2-N-3014 (ref. 32). 845 bp from *EF1a* were amplified using the primer pairs M44-1/rcM53-2, rcm4/M52.7 and ef44/rcM52.6 (refs 33, 34). We also used an internal arctiid-specific primer pair developed from an alignment of a subset of the 26 species: Internal Forward (5' ACGTCTTACGTTGAAACCAAC 3')/Internal Reverse (5' GGACACAGAGATTTCAATGAACAT 3') and a noctuid-specific primer pair developed from a consensus alignment of 18 noctuid species sequenced in ref. 35: Noctuidae Forward (5' TTCGAGAAGGARGCCAG 3')/Noctuidae Reverse (5' GAGGGAAYTCYTGGAGGA 3'). We were only able to obtain a portion of the 845 bp sequence for *C. tenera* and *Euchaetes egle*, so these species were excluded from the final alignment for this locus. 457 bp from *wingless* were amplified using the published primer pairs LepWG1/LepWG2 and LepWG2/LepWG2a³⁶. We also designed internal arctiid-specific primers using the alignment of several of the 26 species: WGIntF (5' TGGTCTGGATTATGAGG CCGCA 3') paired with LepWG1 and WGIntR (5' TCTGGCTCGTGC ACGGTTAAGACC 3') paired with LepWG2. We were unable to amplify *E. egle* for this locus, so this species was not included in the final *wingless* alignment. *L.*

dispar was selected as the outgroup for this analysis because this species is closely related to the arctiids³⁵ and has been used as the outgroup in previous arctiid phylogenies³⁷.

PCR amplification was performed using conditions available from the authors. PCR products were cleaned using Exonuclease I and Shrimp Antarctic Phosphatase, and then purified on Sephadex columns (Sigma-Aldrich). The purified products were sequenced with a Big Dye Terminator Cycle sequencing kit and an ABI-3100 automated sequencer (Applied Biosystems) with the same primers used for amplification. The program Aligner (CodonCode Corporation) was used to edit and align the sequences.

MODELTEST 3.7 (ref. 38) was used to determine the best-fit model of nucleotide substitution for each locus. Using an AIC (Akaike information criterion) approach, which measures the fit of various nucleotide substitution models to the data, the best-fit model was GTR + I + G (where GTR is General Time Reversible, I is the proportion of invariable sites, and G is the shape parameter of the gamma distribution) for mitochondrial *COI*, SYM + I + G for *EF1a* and *wingless*, and GTR + I + G for the three loci combined. The GTR + I + G model was used in the combined maximum likelihood analysis of the three loci in PAUP* 4.0 (ref. 39). The tree resulting from this analysis (not shown) has a nearly identical topology to the 50% majority consensus tree from the bayesian analysis.

MrBayes 3.1.2 allows the user to apply the best-fit model of nucleotide substitution to each locus separately in a combined analysis. As determined from MODELTEST 3.7 above, the GTR + I + G model was applied to the mitochondrial *COI* locus and the SYM + I + G model to the *EF1a* and *wingless* loci. The analysis ran for 10-million generations, with sampling every 1,000 generations. The average value of the potential scale reduction factors was 1.00 and the average standard deviation of split frequencies at the end of the run was 0.001, demonstrating convergence. The first 2,000 trees were eliminated as burn-in, and a 50% majority-rule consensus tree was created using PAUP*4.0. The MrBayes runs were performed at Cornell's Computational Biology Service Unit.

For the comparative analysis using this bayesian tree, we used phylogenetically transformed dummy variables to allow for statistical analyses using both continuous (emergence date and percentage nocturnality) and discontinuous categorical (visual and acoustic signals) traits⁴⁰.

- Stevens, M., Parrage, C. A., Cuthill, I. C., Partridge, J. C. & Troschianko, T. S. Using digital photography to study animal coloration. *Biol. J. Linn. Soc.* **90**, 211–237 (2007).
- Simon, C., Frati, F., Beckenbach, A., Liu, H. & Flook, P. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Ent. Soc. Am.* **87**, 651–701 (1994).
- Cho, S. et al. A highly conserved nuclear gene for low-level phylogenetics: elongation factor 1- α recovers morphology-based tree for heliothine moths. *Mol. Biol. Evol.* **12**, 650–656 (1995).
- Monteiro, A. & Pierce, N. Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from *COI*, *COII*, and *EF1- α* gene sequences. *Mol. Phylogenet. Evol.* **18**, 264–281 (2001).
- Mitchell, A., Mitter, C. & Regier, J. C. Systematics and evolution of the cutworm moths (Lepidoptera: Noctuidae): evidence from two protein-coding nuclear genes. *Syst. Ent.* **31**, 21–46 (2006).
- Brower, A. & DeSalle, R. Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of *wingless* as a source of characters for phylogenetic inference. *Insect Mol. Biol.* **7**, 73–82 (1998).
- Jacobson, N. L. & Weller, S. J. *A cladistic Study of the Arctiidae (Lepidoptera) by Using Characters of Immatures and Adults* (Thomas Say Publications in Entomology, Ent. Soc. Am., Lanham, MD, 2002).
- Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
- Swofford, D. L. *PAUP*: Phylogenetic Analysis using Parsimony (* and other Methods)* v. 4.0b 10 (Sinauer Associates, Sunderland, 2003).
- Dunn, P. O., Whittingham, L. A. & Pitcher, T. E. Mating systems, sperm competition and the evolution of sexual dimorphism in birds. *Evolution* **55**, 161–175 (2001).

LETTERS

The virophage as a unique parasite of the giant mimivirus

Bernard La Scola^{1*}, Christelle Desnues^{1*}, Isabelle Pagnier¹, Catherine Robert¹, Lina Barrassi¹, Ghislain Fournous¹, Michèle Merchat², Marie Suzan-Monti¹, Patrick Forterre^{3,4}, Eugene Koonin⁵ & Didier Raoult¹

Viruses are obligate parasites of Eukarya, Archaea and Bacteria. *Acanthamoeba polyphaga* mimivirus (APMV) is the largest known virus; it grows only in amoeba and is visible under the optical microscope. Mimivirus possesses a 1,185-kilobase double-stranded linear chromosome whose coding capacity is greater than that of numerous bacteria and archaea^{1–3}. Here we describe an icosahedral small virus, Sputnik, 50 nm in size, found associated with a new strain of APMV. Sputnik cannot multiply in *Acanthamoeba castellanii* but grows rapidly, after an eclipse phase, in the giant virus factory found in amoebae co-infected with APMV⁴. Sputnik growth is deleterious to APMV and results in the production of abortive forms and abnormal capsid assembly of the host virus. The Sputnik genome is an 18.343-kilobase circular double-stranded DNA and contains genes that are linked to viruses infecting each of the three domains of life Eukarya, Archaea and Bacteria. Of the 21 predicted protein-coding genes, eight encode proteins with detectable homologues, including three proteins apparently derived from APMV, a homologue of an archaeal virus integrase, a predicted primase–helicase, a packaging ATPase with homologues in bacteriophages and eukaryotic viruses, a distant homologue of bacterial insertion sequence transposase DNA-binding subunit, and a Zn-ribbon protein. The closest homologues of the last four of these proteins were detected in the Global Ocean Survey environmental data set⁵, suggesting that Sputnik represents a currently unknown family of viruses. Considering its functional analogy with bacteriophages, we classify this virus as a virophage. The virophage could be a vehicle mediating lateral gene transfer between giant viruses.

The original strain of APMV, mimivirus, was obtained from a cooling tower in Bradford, UK. Its size challenged the definition of a virus⁶ and led to the idea that giant viruses might be an uncharacterized but important part of the biosphere. We isolated a new strain of APMV, by inoculating *A. polyphaga* with water from a cooling tower, in Paris. We denoted this new strain mamavirus because it seemed to be even larger than mimivirus² when observed by transmission electron microscopy. The main features of mamavirus closely resembled those described for mimivirus, including the formation of a giant viral factory and the typical particle morphology with a multilayered membrane covered with fibrils⁴. We also observed unknown icosahedral small viral particles, 50 nm in size, in virus factories and in the cytoplasm of the infected cells (Fig. 1). Considering the association of this newly detected virus with mamavirus, we named it Sputnik.

Sputnik did not multiply when inoculated into *A. castellanii* (Supplementary Information and Supplementary Table 4).

However, this virus did grow, as demonstrated by transmission electron microscopy and polymerase chain reaction, in *A. castellanii* co-infected with mimivirus or mamavirus (Supplementary Information

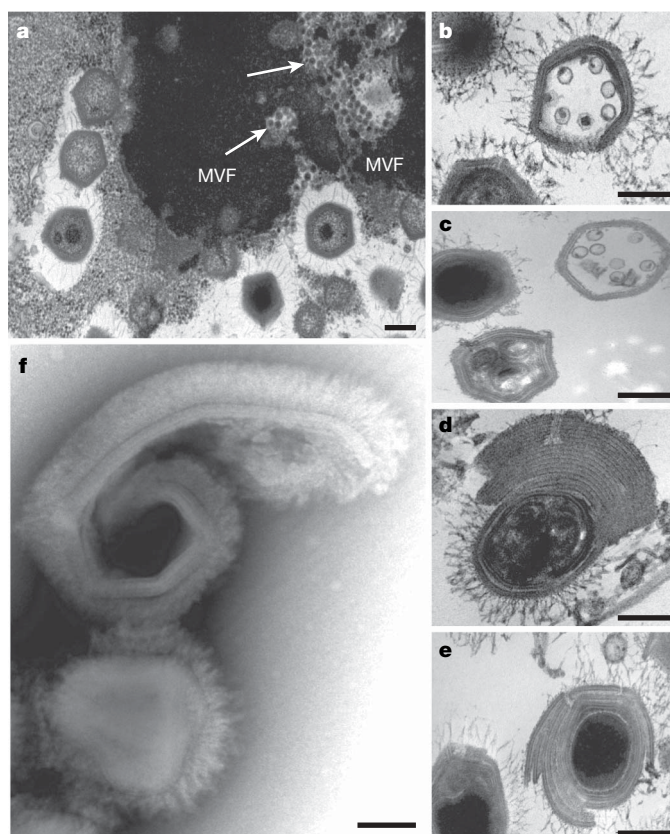


Figure 1 | Different morphological aspects of mamavirus and Sputnik.

a–e, Observations by transmission electron microscopy; **f**, observation by negative staining electron microscopy. **a**, Mamavirus virus factory (MVF) with mamavirus particles at different stages of maturation. Clumps of Sputnik particles (arrows) are observed within MVF. **b**, In some cases, Sputnik is observed within mamavirus capsids. **c**, Defective particles are produced. **d–f**, Co-infection with mamavirus and Sputnik results in abnormal morphology of mamavirus particles, such as membrane accumulation at one side (**d**), membrane accumulation around the particles (**e**), or open particles (**f**). Scale bars, 200 nm.

¹URMITE, Centre National de la Recherche Scientifique UMR IRD 6236, Faculté de Médecine, Université de la Méditerranée, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 5, France. ²Climespace, 185 Rue de Bercy, 75012 Paris, France. ³Biologie Moléculaire du Gène chez les Extrémophiles, Institut de Génétique et Microbiologie, Bâtiment 409, Université Paris Sud, Centre d'Orsay, 91405 Orsay Cedex, France. ⁴Biologie Moléculaire du Gène chez les Extrémophiles, Département de Microbiologie, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France. ⁵National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Building 38A, Room 5N503, 8600 Rockville Pike, Bethesda, Maryland 20894, USA.

*These authors contributed equally to this work.

and Supplementary Table 4). Sputnik and mamavirus were produced within the same viral factory with different kinetics and at different specific locations. Sputnik was produced earlier than APMV (Fig. 2). Sputnik co-infection was associated with a significant increase in the formation of abnormal mamavirus virions, characterized by partial thickening of the capsid (11% rather than 1%, $P = 0.0029$). In the regular mamavirus virions, the capsid layer was 40 nm thick; in contrast, in the presence of Sputnik, the thickness of the capsid wall could reach 240 nm (Fig. 1). In most cases, several capsid layers accumulated asymmetrically at one pole of the viral particle. Some of these abnormal particles seemed to be mature and to harbour fibrils only on the normal part of the capsid. Only a small fraction of the mamavirus particles encapsidated Sputnik virions (Fig. 1). However, co-inoculation of mamavirus with Sputnik resulted in a roughly 70% decrease in the yield of infective mamavirus particles and a threefold decrease in amoeba lysis at 24 h. These findings showed that Sputnik is a parasite of mamavirus that substantially affects the reproduction of the host virus.

The *Acanthamoeba castellanii* mamavirus genome (C.D., B.L.S., C.R., G.F. and D.R., unpublished observations) is about 1,200 kilobase

pairs in size. Its genome is highly AT-rich (A + T content $\approx 72\%$). Orthologues to mimivirus open reading frames (ORFs) were detected for 99% of the predicted mamavirus genes, with amino-acid identity ranging from 75% to 100%. Thus, mamavirus is closely related to mimivirus and could be considered a second strain of APMV. Sputnik has an 18,343-base-pair (bp) circular double-stranded DNA genome, with 21 predicted protein-coding genes ranging in size from 88 to 779 amino-acid residues (Table 1 and Fig. 3). The organization of the Sputnik genome is typical of viral genomes, namely a tight arrangement but little overlap of the ORFs. The high A + T content (73%) of the Sputnik genome is similar to that of APMV. Sputnik samples were resolved by two-dimensional gel electrophoresis within a pI range of 3–10 (Fig. 3). The most abundant of the detected protein spots, analysed by matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) mass spectrometry, corresponded to ORF 20; ORF 08 and ORF 19 proteins were identified once each. These results were corroborated by western blot analysis with a mouse anti-serum against purified Sputnik (Supplementary Fig. 1). Thus, ORF 20 most probably encodes the major capsid protein of Sputnik, whereas ORFs 08 and 19 encode minor virion proteins.

Genomes of many viruses contain a high proportion of ‘ORFan’ genes; that is, genes without detectable homologues in current sequence databases. The genome of Sputnik is no exception because most of its encoded proteins (13 of 21) are ORFans. The eight non-ORFan proteins have viral/plasmid, bacterial or eukaryotic homologues, and/or homologues from the environmental Global Ocean Survey (GOS) data set (Table 1). Three of the Sputnik predicted proteins (ORFs 6, 12 and 13) were most closely related to mimivirus/mamavirus gene products. The proteins encoded in ORFs 12 and 13 were equally similar to their respective homologues from the mimivirus and the mamavirus (Supplementary Table 3), whereas ORF 6 was more closely related to the mamavirus homologue. The most plausible model is therefore that Sputnik acquired a portion of the gene (or the complete gene, which was further partly eliminated) from mamavirus after its divergence from the common ancestor with mimivirus.

Specifically, ORF 12 is uncharacterized, whereas ORFs 6 and 7 encode paralogous proteins containing highly conserved collagen triple-helix motifs⁷. The protein encoded by ORF 13 consists of two domains implicated in viral DNA replication. The carboxy-terminal domain of this protein is a superfamily 3 helicase that is highly conserved and clusters with homologues from nucleocytoplasmic large DNA viruses (NCLDVs)⁷ in phylogenetic trees (Fig. 3 and Supplementary Figs 2 and 3). The amino-terminal portion of ORF 13 protein is a previously unobserved domain for which homologues with high similarity were detected only among proteins from the GOS data set and which, on the basis of the presence of a signature sequence motif, could be predicted to represent a highly derived version of the archaeo-eukaryotic primase (Supplementary Fig. 4). The ORF 3 protein showed limited similarity to a packaging ATPase of the FtsK–HerA superfamily that is found in all NCLDVs and many bacteriophages^{5,8,9} (Fig. 3 and Supplementary Fig. 5). ORF 14, which is adjacent to the primase–helicase gene, encodes a protein containing a Zn-ribbon motif that is significantly similar to that in several proteins in the GOS data set (Table 1 and Supplementary Fig. 6), and ORF 4 also encodes a Zn-ribbon protein without highly conserved homologues. ORF 17 encodes a protein with homologues in the GOS data set that belong to the family of bacterial insertion sequence transposase DNA-binding subunits/domains (transposase A proteins) (Table 1, Fig. 3 and Supplementary Fig. 7). Finally, ORF 10 protein showed significant sequence similarity to integrases of the tyrosine recombinase family from archaeal viruses and proviruses, a relationship that was further supported by phylogenetic analysis (Fig. 3 and Supplementary Fig. 8).

Two genes implicated in essential functions in viral genome replication and packaging (ORFs 13 and 3, respectively) and a gene with a potential role in expression regulation (ORF 14) are most closely

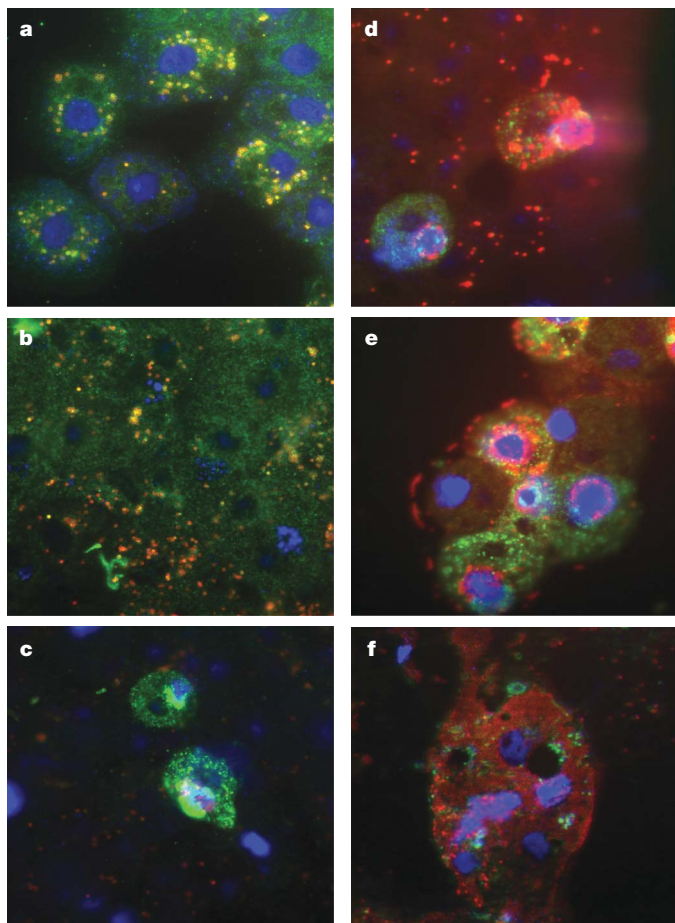


Figure 2 | Sputnik propagation in mamavirus-infected amoebae. *A. castellanii* cells were infected with a mixture of mamavirus and Sputnik. Indirect immunofluorescence labelling was performed with rabbit anti-mimivirus serum (red) and mouse anti-Sputnik serum (green), and nucleic acids were stained with 4,6-diamidino-2-phenylindole (DAPI; blue). **a**, Numerous Sputnik virions entered the cytoplasm at 30 min after infection. **b**, At 4 h after infection, the first viral factories were seen as distinct, strongly stained patches. No viral particles could be seen in these cells, indicating an eclipse phase. **c**, At 6 h after infection, the viral factories expanded and were homogeneously and strongly stained with DAPI. Sputnik production was detected at one side of the viral factory, but no mamavirus virions. **d–f**, At 8 h after infection (**d**), mamavirus production was observed; this increased extensively at 12 h (**e**) and 16 h (**f**) after infection.

related to genes from the GOS data set. Given that the primase–helicase and the FtsK-like ATPase are typical viral genes, it seems likely that Sputnik is linked to an unknown family of viruses, perhaps related to NCLDV, that is abundantly represented among the marine metagenomic sequences but not in other current sequence databases.

Thus, the Sputnik genome contains genes evolutionarily related to at least three distinct sources: first, a putative novel family of viruses; second, an archaeal virus (or plasmid); and third, mimivirus/mamavirus. The three genes shared with mimivirus/mamavirus were probably acquired by Sputnik after the association with APMV was established, and their products might be involved in the interaction of the virophage with its viral host. Within viral factories, recombination between the genomes of the virophage and APMV could result in an exchange of genes. APMV factories are probably capable of replicating foreign DNA, as suggested by experiments demonstrating

efficient plasmid replication in poxvirus¹⁰ and in African swine fever virus factories¹¹. The presence of three genes homologous to mamavirus genes in the Sputnik genome suggests that gene transfer between Sputnik and mamavirus can occur during infection of *Acanthamoeba* by these two viruses together. It has been shown that some bacterial genes were recently acquired by mimivirus¹², but the source and the route of acquisition are still unknown¹³. Virophage could be a vehicle of such gene transfers, as well as of gene transfers between different giant viruses especially, if proviophages exist—a possibility that seems particularly plausible given the presence of genes for the predicted integrase and transposase subunit homologue in the virophage genome.

The integrase gene that is shared between Sputnik and archaeal viruses (plasmids) might have been independently derived from an ancestral virus that predated the divergence between archaea and

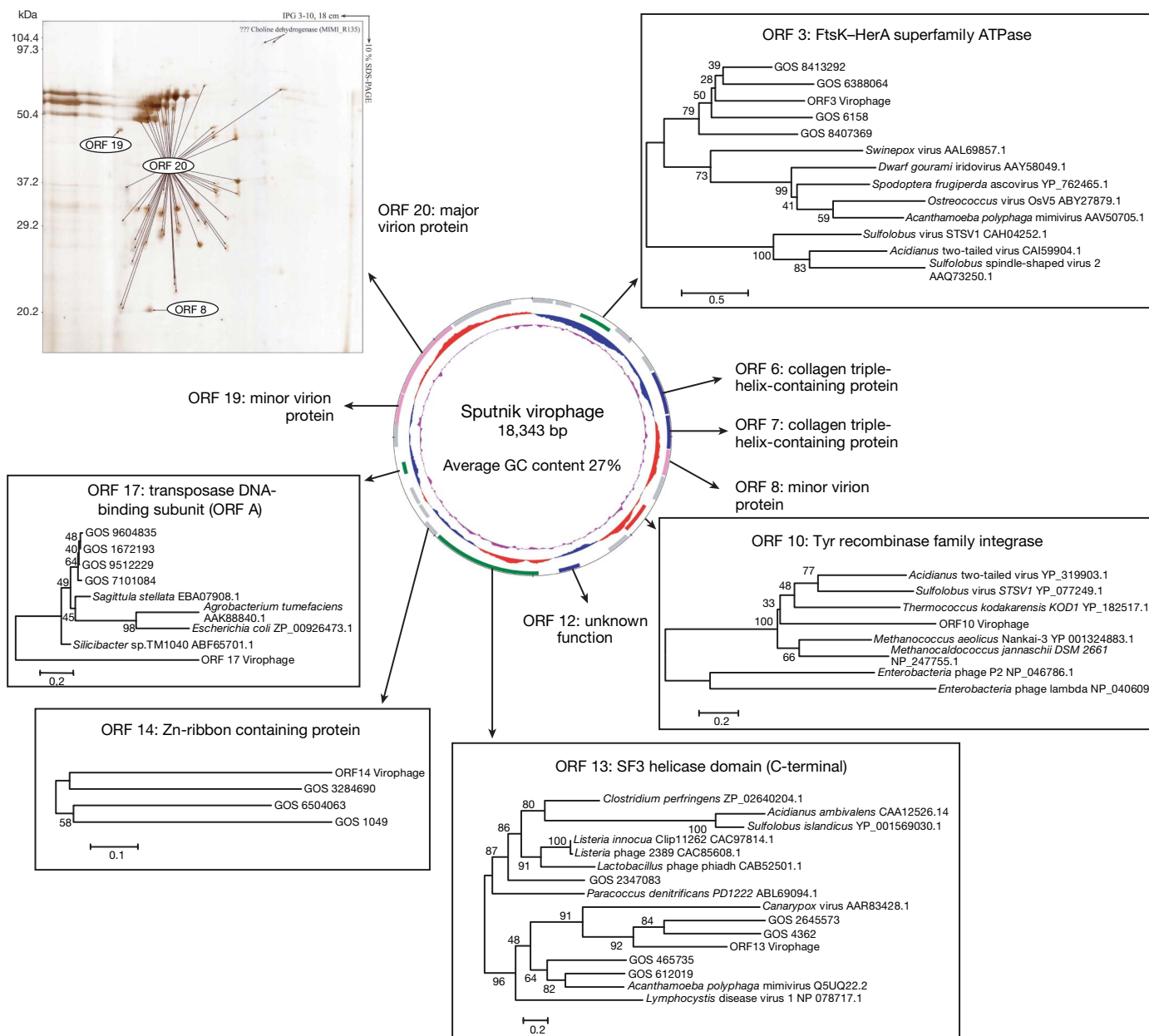


Figure 3 | The Sputnik chromosome. The predicted protein coding sequences are indicated on the two DNA strands (first, outer, circle) and coloured according to their corresponding homologues. ORFs with homologues to mamavirus/mimivirus are indicated in blue, ORFs with homologues to other NCLDVs and bacteriophages are shown in green, and the ORF homologous to an archaeal virus gene is shown in red. The virion

protein coding sequences are shown in purple and ORFans are shown in grey. Phylogenetic trees are displayed for the predicted protein coding sequences with homologues in nr and/or the GOS data sets along with the 2D-gel identifying the capsid protein. GC skew and G + C content are indicated in the second and third circles, respectively. IPG, immobile pH-gradient buffer.

Table 1 | Homologies and predicted functions of the Sputnik protein coding sequences

Gene (size, amino-acid residues)	Closest homologue in GenBank nr (accession no., percentage identity/alignment length/E-value)	Closest homologue in the GOS data set (percentage identity/alignment length/E-value)	Domain architecture/protein family/predicted activity	Predicted function in viroplasm replication
ORF 1 (144)	–	–	Unknown	Unknown
ORF 2 (114)	–	–	Unknown	Unknown
ORF 3 (245)	RecA-superfamily ATPases (<i>Actinobacillus pleuropneumoniae</i> serovar 1 str. 4074) (ZP_00134596.2, 54%/35/0.01) MIMI L712	GOS_6857935 (48%/205/10 ^{−37})	FtsK–HerA superfamily ATPase	DNA packaging
ORF 4 (139)	Limited similarity to diverse Zn-ribbon proteins	–	Zn-ribbon-containing protein	Transcription regulation?
ORF 5 (119)	–	–	Unknown	Unknown
ORF 6 (310)	MIMI R196 (YP_142550.1, 53%/128/4 × 10 ^{−19})	GOS_3129237 (59%/130/10 ^{−23})	Collagen triple-helix-repeat-containing protein	Protein–protein interactions in factories
ORF 7 (236)	C1q and tumour necrosis factor related protein 5, mouse (NP_663588, 27%/156/0.001) MIMI R239	GOS_8448924 (57%/40/0.002)	Collagen triple-helix-repeat-containing protein	Protein–protein interactions in factories
ORF 8 (184)	–	–	Unknown	Minor virion protein
ORF 9 (175)	–	–	Unknown	Unknown
ORF 10 (226)	Phage integrase family protein (<i>Methanococcus aeolicus</i> Nankai-3) (YP_001324883, 32%/166/6 × 10 ^{−13})	–	Tyr recombinase family integrase	Integration of viroplasm into APMV genome?
ORF 11 (162)	–	–	Unknown	Unknown
ORF 12 (152)	MIMI R546 (Q5UR26, 64%/122/5 × 10 ^{−42})	–	Unknown	Unknown
ORF 13 (779)	Putative DNA-polymerase or DNA-primase (<i>Lactobacillus phage phiadh</i>) (NP_050131.1, 29%/171/4 × 10 ^{−12}) MIMI L207/206	Putative highly derived N-terminal primase domain, GOS_5022207 (32%/200/8 × 10 ^{−18}) C-terminal SF3 helicase domain GOS_2645573 (32%/409/4 × 10 ^{−46})	Primase–helicase	DNA replication
ORF 14 (114)	–	GOS_3284690 (45%/48/0.02)	Zn-ribbon-containing protein	Transcription regulation?
ORF 15 (109)	–	–	Membrane protein	Modification of APMV membrane?
ORF 16 (130)	–	–	Unknown	Unknown
ORF 17 (88)	–	GOS_9512229 (27%/80/0.03)	IS3 family transposase	DNA-binding protein
ORF 18 (167)	–	–	A protein	Unknown
ORF 19 (218)	–	–	Unknown	Minor virion protein
ORF 20 (595)	–	–	Unknown	Major capsid protein
ORF 21 (438)	–	–	Unknown	Unknown

eukaryotes. Alternatively, Sputnik might have acquired this gene from a virus (plasmid) harboured by an archaeal endosymbiont residing in a eukaryotic cell infected by Sputnik. Regardless of the exact source of this gene, one of the most remarkable features of the viroplasm is its apparent chimaeric origin. This seems to be one of the most convincing cases so far of gene mixing and matching within the virus world¹⁴. A search for additional viroplasm should shed more light on this unique mode of interaction between viruses.

As Sputnik multiplies in the APMV giant factories, it resembles satellite viruses of animals (for example adeno-associated viruses or hepatitis D virus) and plants (for example satellite tobacco necrosis virus)¹⁵. However, Sputnik reproduction seems to impair the production of normal APMV virions significantly, indicating that it is a genuine parasite. To our knowledge, this observation of a virus using the viral factory of another virus to propagate at the expense of its viral host has not been described previously. We have therefore termed this virus a viroplasm by analogy with bacteriophages; should other similar agents be discovered in the future, viroplasm could be used as a generic name to denote them.

METHODS SUMMARY

Isolation of viruses was performed on water sampled in a cooling tower as described previously¹⁶. For developmental cycle analysis, *A. castellanii* cells were infected with mamavirus alone or with Sputnik (Supplementary Information) and examined by transmission electronic microscopy and fluorescence as described previously for mimivirus¹.

Large volumes of *A. castellanii* infected by mamavirus and Sputnik were cultured. The culture supernatants were then filtered through 0.8-μm and 0.2-μm membranes. Sputnik particles were concentrated from the 0.2-μm filtrate, whereas mamavirus was obtained by washing the 0.2-μm membranes with K36 buffer. DNA was extracted by following the mimivirus procedure¹. The genomes of the two viruses were sequenced on the 454-Roche GS20 as described¹⁷. Putative

open reading frames were searched with GeneMark.hmm 2.0 (ref. 18), and translated sequences were compared with GenBank nr and the GOS data set (<http://www.ncbi.nlm.nih.gov>). MAFFT version 6 (ref. 19) or MUSCLE²⁰ was used to construct multiple alignments, and MEGA 4 (ref. 21) or TREEFINDER²² was used to construct phylogenetic trees. Peptide data from excised spots were analysed by MALDI-TOF mass spectrometry as reported previously²³. For western blot analysis, sera of BALB/c mice immunized with mamavirus or Sputnik were first absorbed on mimivirus and then on amoebae lysate.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 June; accepted 27 June 2008.

Published online 6 August 2008.

1. Raoult, D. *et al.* The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344–1350 (2004).
2. La Scola, B. *et al.* A giant virus in amoebae. *Science* **299**, 2033 (2003).
3. Koonin, E. V. Virology: Gulliver among the Lilliputians. *Curr. Biol.* **15**, R167–R169 (2005).
4. Suzan-Monti, M., La Scola, B., Barrassi, L., Espinosa, L. & Raoult, D. Ultrastructural characterization of the giant volcano-like virus factory of *Acanthamoeba polyphaga* Mimivirus. *PLoS ONE* **2**, e328 (2007).
5. Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
6. Raoult, D. & Forterre, P. Redefining viruses: lessons from Mimivirus. *Nature Rev. Microbiol.* **6**, 315–319 (2008).
7. Rasmussen, M., Jacobsson, M. & Björck, L. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* **278**, 32313–32316 (2003).
8. Williamson, S. J. *et al.* The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* **3**, e1456 (2008).
9. Iyer, L. M., Makarova, K. S., Koonin, E. V. & Aravind, L. Comparative genomics of the FtsK–HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res.* **32**, 5260–5279 (2004).

10. De Silva, F. S. & Moss, B. Origin-independent plasmid replication occurs in vaccinia virus cytoplasmic factories and requires all five known poxvirus replication factors. *Viral. J.* **2**, doi:10.1186/1743-422X-2-23 (2005).
11. Oliveira, S. & Costa, J. V. Replication of transfected plasmid DNA by cells infected with African swine fever virus. *Virology* **207**, 392–399 (1995).
12. Filee, J., Siguier, P. & Chandler, M. I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *Trends Genet.* **23**, 10–15 (2007).
13. Moreira, D. & Brochier-Armanet, C. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. *BMC Evol. Biol.* **8**, doi:10.1186/1471-2148-8-12 (2008).
14. Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient Virus World and evolution of cells. *Biol. Direct* **1**, doi:10.1186/1745-6150-1-29 (2006).
15. Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. (eds) *Virus Taxonomy (Eighth Report of the International Committee on Taxonomy of Viruses)* 1163–1169 (Elsevier, London, 2005).
16. La Scola, B., Barrassi, L. & Raoult, D. Isolation of new fastidious α -Proteobacteria and *Afipia felis* from hospital water supplies by direct plating and amoebal co-culture procedures. *FEMS Microbiol. Ecol.* **34**, 129–137 (2000).
17. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
18. Lukashin, A. V. & Borodovsky, M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115 (1998).
19. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
20. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
21. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
22. Jobb, G., von Haeseler, A. & Strimmer, K. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol. Biol.* **4**, doi:10.1186/1471-2148-4-18 (2004).
23. Kowalczywska, M. & Raoult, D. Advances in *Tropheryma whippelii* research: the rush to find biomarkers for Whipple's disease. *Future Microbiol.* **2**, 631–642 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank X. de Lamballerie, S. Azza, P. de Clocquement, L. Espinosa, B. Campagna, N. Aldrovandi, V. Brice, A. Bernard, C. Ivars, B. Giumelli and Y. Wolf for expert assistance. This work was funded by the Centre National de la Recherche Scientifique (CNRS, crédits récurrents). I.P. is funded by a CIFFRE fellowship, E.K. is supported by the Intramural Research Program of the National Institutes of Health, National Library of Medicine, and P.F. is funded by the Institut Universitaire de France.

Author contributions D.R. and B.L.S. supervised the project and wrote the manuscript. C.D., P.F. and E.K. contributed to sequence analysis, interpretation of the results and writing of the manuscript. I.P. isolated the virus. M.S.-M. contributed to viral cycle analysis, interpretation of the results and writing of the manuscript. M.M. provided water samples. L.B. conducted the viral cycle experiment. C.R. and G.F. sequenced the genome.

Author Information The virophage genome has been deposited in GenBank under accession number EU606015. The *Acanthamoeba castellanii* *mamavirus* genes with homologues found in the Sputnik genome have been deposited in GenBank under accession numbers EU827539–EU827541. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R. (didier.raoult@gmail.com).

METHODS

Inactivation of Sputnik. To obtain a pure suspension of mamavirus we proposed that, as observed previously for mimivirus¹, it would be resistant to high temperatures. We therefore subjected a supernatant containing Sputnik and mamavirus to 65 °C for 1 h. This suspension was then diluted in PAS (Page's amoebal saline) buffer by tenfold dilutions from 10⁻¹ to 10⁻¹⁰. Each dilution was inoculated into four culture wells of a suspension of fresh amoebae and observed daily for lysis under an inverted microscope. The last dilution producing lysis in one in four wells was 10⁻⁵. The supernatant of this well was subcultured onto fresh amoebae, and an absence of Sputnik was verified by transmission electronic microscopy, immunofluorescence staining and Sputnik-specific PCR (see Supplementary Methods and Supplementary Results).

Evaluation of the effect of Sputnik on the developmental cycle of mamavirus.

Supernatant containing Sputnik and mamavirus from infected *A. castellanii* was filtered through a 0.2-µm membrane and the Sputnik-containing filtrate was saved. A suspension of 10 ml of pure mamavirus was divided between two tubes. In tube 1, 200 µl of the Sputnik-containing supernatant was added. In tube 2, 200 µl of PAS buffer was added. *A. castellanii* cells (10 ml, 5 × 10⁵ ml⁻¹ in PAS buffer) were inoculated into culture flasks. In one flask, 1 ml of tube 1 was added; in a second flask, 1 ml of tube 2 was added, and 1 ml of PAS was added in the third flask. Living trophozoites were counted in each flask after 24 h. At 48 h after inoculation, mamavirus (flask 2) or Sputnik and mamavirus (flask 1) culture supernatants were used for titration of mamavirus and were then frozen. Titration was performed by endpoint dilution from 10⁻¹ to 10⁻¹⁰ as described above and then with fivefold dilutions from 10⁻⁴ to 10⁻⁶. Dilutions were scored until day 5 for lysis indicating mamavirus multiplication. The presence or absence of mamavirus multiplication was confirmed by detection with PCR in the supernatants from wells (data not shown).

To evaluate the effect of Sputnik on the appearance of abnormal mamavirus particles, monolayers of *A. castellanii* cells infected by mamavirus alone and by Sputnik and mamavirus were prepared for transmission electron microscopy. To normalize the comparison, counts of viral particles were performed in an area with a width of 1.5 µm around the virus factory limits.

Purification of viruses, preparation of viral DNA, and sequencing of Sputnik virus and mamavirus genomic DNA. Large volumes of *A. castellanii* cells infected by mamavirus and Sputnik were cultured. Viral supernatant were collected at 24–48 h, when lysis of amoebae was almost complete, by low-speed (100g) centrifugation for 15 min.

Sputnik was purified by filtration on 0.8-µm and 0.2-µm membranes. The filtrate was concentrated by ultracentrifugation at 100,000g for 70 min at 4 °C. The pellet was resuspended in K36 buffer, loaded on a 25% sucrose cushion in K36 and centrifuged with the same conditions. The purified pellet was washed once in K36 and resuspended in 10 mM Tris-HCl, 1 mM EDTA. To avoid contamination from DNA and RNA from amoebae, the suspension was treated twice with 10 µl of DNaseI_RNaseI-free (Roche) and 10 µl of RNaseI_DNaseI-free (Roche) and incubated for 60 min at 37 °C. The enzymes were inactivated by

heating for 10 min at 70 °C. The DNA was extracted by following the mimivirus procedure¹. A semiquantitative PCR was performed with primers specific for the 18S rRNA gene from amoebae¹ to estimate the contamination with DNA from amoebae. The Sputnik genome was pyrosequenced on 454–Roche GS20 as described¹⁷. The raw data were assembled by the gsAssembler of the GSFLX (35-bp overlap; 95% identity) leading to a large contig of 16.9 kilobases (kb) and four smaller contigs, for a total of 1.08 kb. Four primer sets were designed to close the molecule by PCR.

To obtain mamavirus DNA, the 0.2-µm membranes were washed with K36 buffer and this suspension was processed as above for sucrose density purification and for treatments with DNase/RNase. The pellet was then resuspended in TSD buffer (40 mM Tris-HCl pH 8, 2% SDS, 60 mM dithiothreitol) and incubated for 30 min at 60 °C with checking for lysis. If needed, an additional 25 µl of buffer was added to achieve total lysis, and this could be repeated three times. The suspension was diluted 1:10 in 50 mM Tris-HCl and treated with 10% Proteinase K at 56 °C. After three phenol/chloroform extractions, the DNA was precipitated with ethanol and resuspended in 75 µl of 10 mM Tris-HCl, 1 mM EDTA. The quality and the yield of the DNA was analysed on an agarose gel and stained with ethidium bromide. A semiquantitative PCR was performed with primers targeting the 18S rRNA gene from amoebae¹ to estimate contamination with DNA from amoebae. The mamavirus genome was also sequenced on 454–Roche GS20 and assembled with gsAssembler (40-bp overlap; 90% identity); 43 large contigs (more than 1.5 kb) were constructed for a genome size of 1.18 megabases. The average contig size was 27 kb; the largest was 173 kb. Taking into account all the contigs, 163 were obtained for a genome size of about 1.20 megabases.

Sequence analyses. Putative ORFs were defined with GeneMark.hmm 2.0 (ref. 18). Significant similarities of the ORF translated sequences were assessed through BLASTP and psi-BLAST²⁴ searches against the NCBI non-redundant protein database (<http://www.ncbi.nlm.nih.gov>). Functional motifs and conserved domains were identified by searches against PFAM version 22.0 (ref. 25), the Conserved Domain Database (CDD version 2.13), and SMART²⁶. Homologues of Sputnik proteins in the environmental sequence data were detected by searching the NCBI environmental data set using BLASTP. Analyses of GC percentages and GC skew were performed with the online DNA Base Composition Analysis Tool (<http://molbiol-tools.ca>). The genome map was generated with Genomeviz²⁷. MAFFT version 6 (ref. 19) or MUSCLE²⁰ was used to construct multiple alignments. Phylogenetic analyses were conducted with MEGA 4 (ref. 21) or TREEFINDER²².

24. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
25. Finn, R. D. *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251 (2006).
26. Letunic, I. *et al.* SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* **32**, D142–D144 (2004).
27. Ghai, R., Hain, T. & Chakraborty, T. GenomeViz: visualizing microbial genomes. *BMC Bioinformatics* **5**, doi:10.1186/1471-2105-5-198 (2004).

Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes

Dacheng Tian^{1*}, Qiang Wang^{1*}, Pengfei Zhang¹, Hitoshi Araki^{1,2}, Sihai Yang¹, Martin Kreitman³, Thomas Nagylaki³, Richard Hudson³, Joy Bergelson^{1,3} & Jian-Qun Chen¹

Mutation hotspots are commonly observed in genomic sequences and certain human disease loci^{1–7}, but general mechanisms for their formation remain elusive^{7–11}. Here we investigate the distribution of single-nucleotide changes around insertions/deletions (indels) in six independent genome comparisons, including primates, rodents, fruitfly, rice and yeast. In each of these genomic comparisons, nucleotide divergence (D) is substantially elevated surrounding indels and decreases monotonically to near-background levels over several hundred bases. D is significantly correlated with both size and abundance of nearby indels. In comparisons of closely related species, derived nucleotide substitutions surrounding indels occur in significantly greater numbers in the lineage containing the indel than in the one containing the ancestral (non-indel) allele; the same holds within species for single-nucleotide mutations surrounding polymorphic indels. We propose that heterozygosity for an indel is mutagenic to surrounding sequences, and use yeast genome-wide polymorphism data to estimate the increase in mutation rate. The consistency of these patterns within and between species suggests that indel-associated substitution is a general mutational mechanism.

Mutation-rate heterogeneity is known to occur at multiple physical scales^{12,13}. Base substitutions and indels positively co-vary both within and between species^{10,14–17}, but this correlation is generally assumed to be indirect, reflecting either the mutability of sequences influenced by their compositional or structural properties, or alternatively the intensity of natural selection acting on both types of mutation in response to a region's functional constraint^{10,16}. This family of indirect causations is henceforth referred to as the 'regional difference' hypothesis. As an alternative, we consider a hypothesis, indel-induced mutation, that causally links indels and single-nucleotide changes: heterozygosity for an indel increases the occurrence of nucleotide changes at nearby sites^{14,18}.

To evaluate the general effect of indels on regional mutation rates, we first investigated nucleotide substitution rates around indels in the following genome-wide comparisons: human (*Homo sapiens*) and chimpanzee (*Pan troglodytes*), human and rhesus macaque (*Macaca mulatta*), mouse (*Mus musculus*) and rat (*Rattus norvegicus*), two rice lines (*Oryza sativa* L. var. Nipponbare versus var. 93-11), and three baker's yeast strains (*Saccharomyces cerevisiae* strain S288C versus RM11-1a and S288C versus YJM789). These genomic comparisons cover a wide range of species and levels of divergence (different genera to same species), and have high-quality sequence alignments (3,318 megabases (Mb)). We carried out extensive tests to convince ourselves that the results are not artefacts of alignment algorithms (see Methods). Our alignments yielded estimates of overall nucleotide diversity or divergence (D) ranging from 0.0045 to

0.1449, and the number of indels per kilobase (kb) ranging from 0.51 to 13.02 among comparisons (Supplementary Table 1), consistent with previous reports^{19,20}.

We plotted the magnitude of D against the distance interval to the nearest indel (d_1 , Fig. 1a) and the length of the indel interval (d_2 , as an index of the reciprocal of indel density, Fig. 1b), both of which are represented by non-overlapping windows ($L_0, R_0, \dots, L_n, R_n$; see Methods and Supplementary Fig. 1). Figure 1 reveals a consistent negative relationship between average D and both distance to the nearest indel (Fig. 1a) and indel interval length (Fig. 1b). Between mouse and rat (top of Fig. 1a), for example, the average D in regions ≤ 100 bp from indels ($d_1 = 0$) is 0.141 but drops to 0.056 in the interval 450–550 bp from indels ($d_1 = 5$). To understand better the effect of indels on D , we analysed the relationship between average D and d_1 for each indel interval (Fig. 2). In comparisons between human and chimpanzee (Fig. 2a) and between two rice lines (Fig. 2b), D is greatest at the distance intervals nearest to the indel, regardless of indel density (d_2 classes), and declines monotonically with distance. Similar patterns are present in the other genome comparisons. In every case, the most rapid decline in D occurs in the first few windows closest to the indel. High indel density augments the effect on D : shorter intervals spanned by two indels (for example, <1 kb) have higher divergence than intervals having the same distance to one nearest indel. Consistent results by other alignment methods (Supplementary Figs 2–4) indicate that the patterns depicted in Figs 1 and 2 are not sensitive to the alignment procedure.

The increase in nucleotide substitution with decreasing distance to an indel is compatible with both the regional difference and the indel-associated mutation hypothesis and does not distinguish between them. We therefore investigated two specific predictions of the indel-associated mutation hypothesis that control for selective constraint. First, in comparing indel differences between two closely related species, the hypothesis predicts a greater number of nucleotide substitutions surrounding an indel on the phylogenetic branch containing the indel mutation than on the branch with the ancestral allele. This prediction can be tested by using an outgroup to infer derived and ancestral states for the indel and surrounding nucleotide differences. Rate differences cannot be attributed to differential selective constraints under this test, because the comparison is between strictly orthologous sequences.

We applied the test to four independent comparisons and closely related outgroups: two races of rice (93-11 versus Nipponbare, *O. nivara* outgroup), human versus chimpanzee (chromosome 1 and 2, rhesus outgroup), human versus chimpanzee (chromosome 7, baboon outgroup), and *Drosophila simulans* versus *D. sechellia* (*D. melanogaster* outgroup). In all four tests (Fig. 3a–d and

¹State Key Laboratory of Pharmaceutical Biotechnology, Department of Biology, Nanjing University, Nanjing 210093, China. ²Department of Fish Ecology and Evolution, EAWAG Center of Ecology, Evolution and Biogeochemistry, 6047 Kastanienbaum, Switzerland. ³Department of Ecology & Evolution, University of Chicago, Chicago, Illinois 60637, USA.

*These authors contributed equally to this work.

Supplementary Table 2) significantly greater numbers of mutations are present on the indel lineage (D_i) than on the non-indel lineage (D_{ni}) in the interval closest to the indel (<150 bp); the significant relative excess in substitutions extended in some instances to more distant intervals. These results are consistent with indel-associated mutation but not with the regional difference hypothesis.

The indel-associated mutation hypothesis makes a similar prediction for indels segregating within a species. We posited that for an indel to be mutagenic to its surrounding sequence, the mutagenic effect must be exerted only while it is segregating as a heterozygote with a non-indel allele. Mechanistically, indel heterozygosity is expected to affect localized chromosome pairing during meiosis and might target the region for mutational repair^{2,10,21}. For example, double-strand DNA breaks have been linked to higher mutation rate²², and template switching during replication is shown to promote errors⁹. In homozygotes, or once the indel is fixed in a species, there is no reason to suppose increased mutagenesis.

Under this model, assuming selective neutrality for both indels and nucleotide mutations, and an equal rate of induction of single-nucleotide mutations on the indel and the non-indel alleles in indel heterozygotes, we obtained an expression for the factor, f , by which the mutation rate increases as a function of the observed quantities N_i and N_{ni} (the numbers of mutations on the derived indel allele and the

ancestral non-indel allele, respectively) for arbitrary sample size and indel allele frequency (see Supplementary Material 2). According to this model, more mutations are expected on the indel allele than the non-indel allele while the indel is segregating.

To test this prediction we analysed indel polymorphisms in three sequenced *S. cerevisiae* strains (S288C, RM11 and YJM789) using the congener *S. paradoxus* as outgroup. Pairwise divergences between the strains are approximately equal ($D = 0.005$), as expected for strains drawn from an unstructured population. Indels can occur at either one-third or two-thirds frequency in this sample, and for these frequencies the ratios of the expected number of mutations on the indel:non-indel alleles are $r = (7f + 17)/(3f + 21)$ and $r = (5f + 9)/(3f + 11)$, respectively, where f is the indel mutation rate increase in heterozygotes (μ_{het}) relative to the background rate (homozygotes, μ_{hom}). We analysed a total of 1,027 indels at one-third frequency and 252 indels at two-thirds frequency. In both cases there are significantly more mutations on indel alleles than non-indel allele (Fig. 3e, f and Supplementary Table 3). Therefore, we proceeded to estimate f . For the one-third frequency indels, the rate increase of μ_{het} to μ_{hom} is statistically significant around the indels ($f = 34.7$ in window 0; $f = 4.2$ in window 1). For the two-thirds frequency indels, the estimated rate increase around the indel is smaller ($f = 4.5$, window 0), but remains nearly constant through the next five windows ($f = 4.61$, windows 1–5). The bootstrap confidence intervals for estimates of f based on the one-third and two-third indel frequency data sets overlap (Supplementary Table 3). Because our model does not incorporate recombination between indel and non-indel chromosomes (or recurrent mutation), the two-thirds frequency indel estimates of f may be more conservative than the corresponding one-third indel frequency estimates.

The yeast population-genetic estimates of f are also compatible with the divergence rate increases we observed between closely

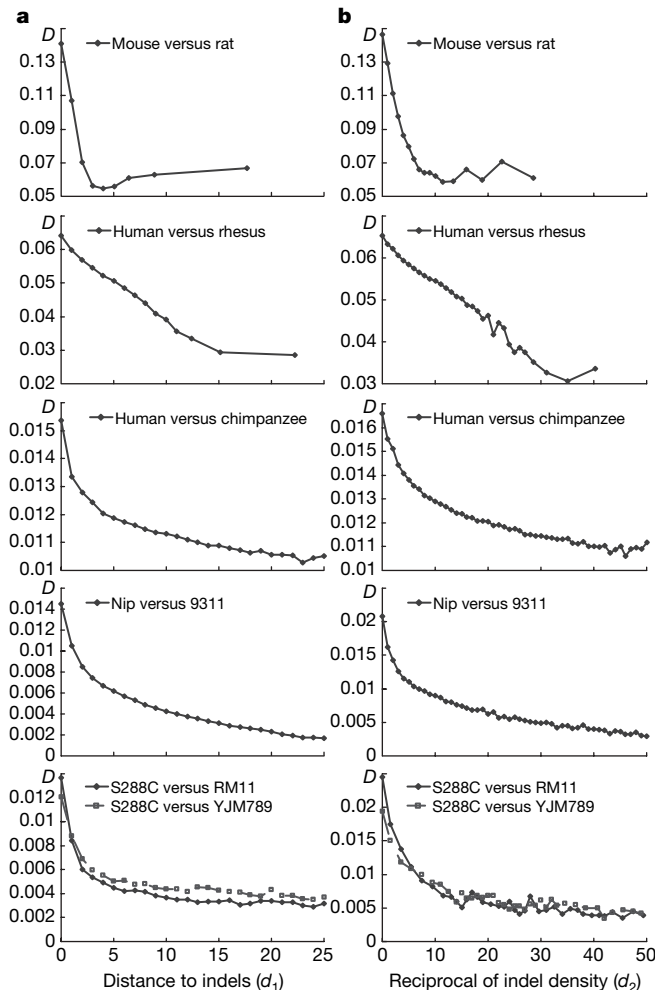


Figure 1 | Nucleotide divergence (D) as a function of (a) distance to nearest indel (d_1) and (b) reciprocal of indel density (d_2). The order from top to bottom is based on the level of divergence (Supplementary Table 1): mouse versus rat, human versus rhesus macaque, human versus chimpanzee, rice subspecies (Nipponbare versus 9311), and yeast strains (SC288 versus RM11 and SC288 versus YJM789). Only the windows $d_1 \leq 25$ or $d_2 \leq 50$ are shown; each data-point contains $>1,000$ samples.

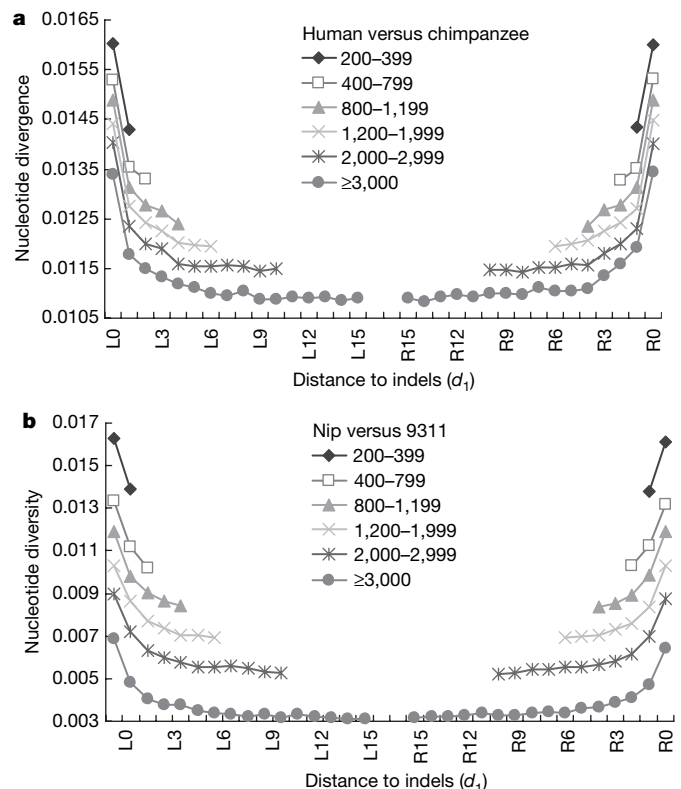


Figure 2 | Relationships between nucleotide divergence (D) and the distance to indels (d_1) as a function of indel interval length (d_2). The lines from top to bottom represent the length of intervals (d_2): 200–399, 400–799, ..., $\geq 3,000$ bp for the comparison between human and chimpanzee (a) and two rice lines (b).

related species. For an indel destined to drift to fixation in the population we can calculate the expected mutation rate surrounding the indel. Conditional on fixation, an indel is expected to spend an equal number of generations at each frequency in the population ($1/2N_e \leq p_{\text{indel}} \leq 1$) and therefore is expected to be in heterozygotes 50% of the time before fixation. The latter implies $E(\mu|\text{indel} \rightarrow \text{fixation}) = (\mu_{\text{het}} + \mu_{\text{hom}})/2$. Assuming, for example, $\mu_{\text{het}} \approx 10\mu_{\text{hom}}$ for the 50 bases surrounding an indel, we should

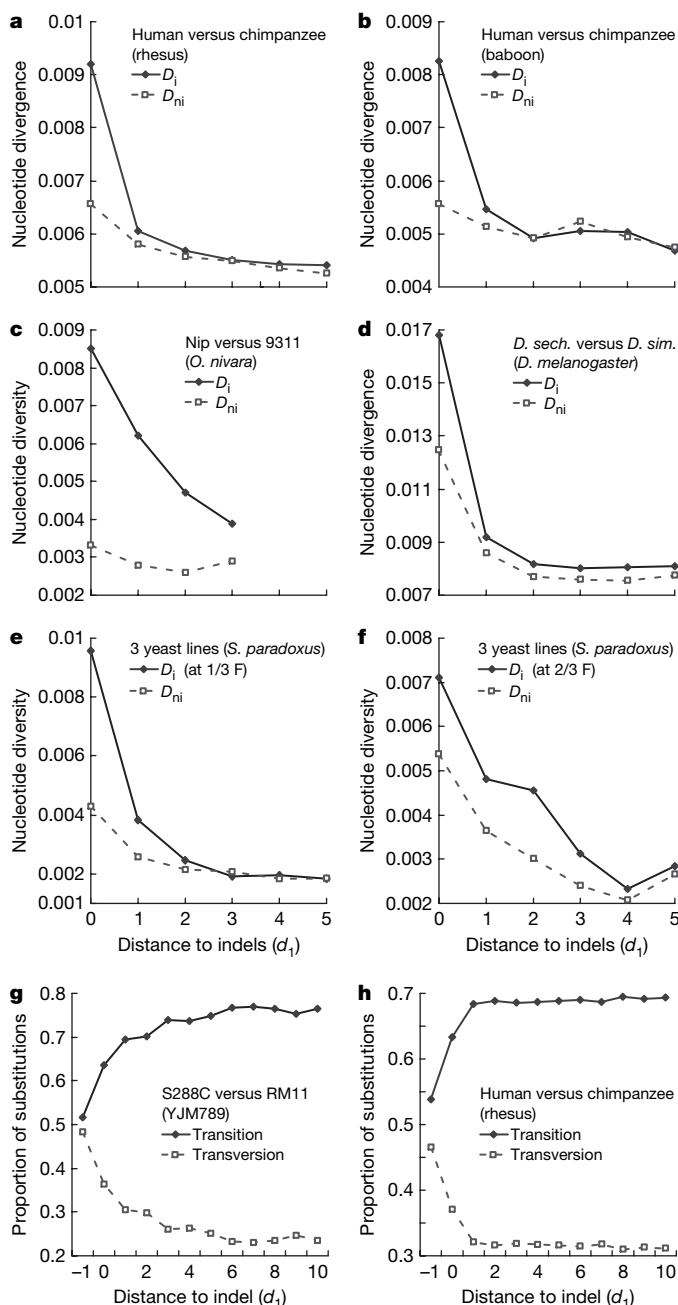


Figure 3 | Three-lineage comparisons. **a, b**, Comparison between human and chimpanzee (rhesus and baboon outgroups, respectively). **c**, Comparison between two rice subspecies (*O. nivara* outgroup). (Not enough data are available for window 4–5.) **d**, Comparison between *D. simulans* and *D. sechellia* (*D. melanogaster* outgroup). **e, f**, Comparison between indel alleles at either one-third or two-thirds frequency and their corresponding non-indel alleles among three yeast strains (*S. paradoxus* outgroup). **g, h**, Transition/transversion rates in relation to d_1 in yeasts (**g**) and human–chimpanzee (**h**). $d_1 = -1$ corresponds to W0'. The lines in **e, f** represent the total proportion of transitions (solid line) or transversions (dashed line), respectively.

expect a fivefold increase in the nucleotide substitution rate immediately surrounding a newly fixed indel, and a lesser increase for indels that have fixed further back in time. For the closely related species comparisons in our data set, the observed increases in divergence very close compared with far from indels are generally consistent with these predictions.

Indel-associated mutations may have a distinctive base substitution signature. One possible signature is the ratio of transitions to transversions, which we find differs from the background in the region immediately surrounding an indel (Fig. 3g, h). The general pattern is an increase in the proportion of transversions ($A \leftrightarrow T$ in particular) relative to transitions. One exception is the pattern of transversion $G \leftrightarrow C$ in mammals, where the proportion of this transversion decreases rather than increases (Supplementary Fig. 5). In addition, the level of nucleotide divergence surrounding an indel is positively correlated with both indel size (Supplementary Fig. 6a–d) and distance to linked indels in the local region (Supplementary Fig. 6e, f) (both by t -test, $P < 0.001$), both of which further support indel-associated mutations. Furthermore, the relationship between distance to an indel (d_1) and divergence (D) is very similar in coding versus non-coding regions (Supplementary Fig. 7), providing an additional indication that indel-associated mutation rather than regional constraint differences are contributing predominantly to the pattern.

Our results are consistent with a model in which a heterozygous indel induces nucleotide mutations in the surrounding DNA. However, the association between indels and mutations does not preclude other possibilities, such as a common cause for indels and mutations, or the induction of the former by the latter. One alternative we considered is that of elevated mutation rates (both indel and nucleotide) in recombination hotspots. We carried out extensive analyses of indels and base substitutions in human–chimpanzee and yeast sequences but found no evidence to support the hypothesis that indel-associated mutation is recombination-rate dependent (Supplementary Material 3). A molecular-level understanding will probably be required to resolve the mechanism driving indel-associated mutation.

Our results indicate that indel-associated mutation occurs throughout the Eukarya. Viewing an indel as a ‘mutator’ has interesting consequences. Indels are generally deleterious in coding regions, and they rarely reach high frequencies in populations. Indeed, indel density in the coding regions of our six genome comparisons occurs at 22.7% relative to non-coding sequences. Although coding indels rarely reach high allele frequencies, they will almost always be in heterozygotes and, according to our model, could contribute disproportionately to base changes in the surrounding coding regions. Non-coding regions can better tolerate indels, and indels will have the greatest impact in these regions. In particular, indel mutation of *cis*-regulatory sequences could contribute to a high rate of evolution for gene expression. Therefore, these ubiquitous mutators may disproportionately contribute to processes linked to speciation and phenotypic evolution. Other recent work has similarly revealed the importance of indels as a source of genetic variation^{23,24}. Our study suggests that the role played by indels in molecular and genome evolution is more important than previously believed.

METHODS SUMMARY

Alignments and analyses. All genome sequences were aligned by Blast²⁵. All alignments analysed satisfied the criteria that indels were < 101 bp and alignments were > 10 kb. Each of the alignments was dissected into non-overlapping windows, each having a specific distance to the nearest indel (d_1). An interval is defined as a region between two indels, the length of which reflects the indel scarcity (d_2).

The windows in an interval are ordered as L0 (the first 50 bp), L1 (100 bp), ..., L_n (and/or R_n , 100 bp), ..., R1 (100 bp), and R0 (the last 50 bp), respectively. An interval 11–99 bp long is assigned to W0'. By definition, L1 or R1 is 0.1 kb closer than L2 or R2 to an indel, respectively, and L_n (and/or R_n) is located at the centre

of an interval. The total length (d_2) of the contiguous, indel-free window is equal to $L1 + \dots Ln + Rn + \dots + R1$, in which the n varies among intervals.

Three-(or four)-lineage comparisons. An outgroup sequence allowed us to identify parsimoniously the lineage in which indel events occurred, and to determine the number of nucleotide substitutions at fixed intervals from the indel in the lineage with the indel (N_i or D_i for divergence) and without the indel (N_{ni} or D_{ni}). For conservative estimates, indels were discarded if the interval was <100 bp and the loci had >1 indels among three sequences or contained slippage-like indels. The ratios $r = N_i/N_{ni}$ or $R = D_i/D_{ni}$ provide an estimate of the relative substitution-rate difference attributable to the indel. In the absence of any association between an indel and the nucleotide substitution rate, the expected number of mutations on indel-containing and non-indel-containing branches are expected to be the same; with indel-associated mutation, we expect $N_i > N_{ni}$ and $D_i > D_{ni}$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 January; accepted 17 June 2008.

Published online 20 July 2008.

- Hedrich, K. *et al.* Distribution, type, and origin of Parkin mutations: review and case studies. *Mov. Disord.* **19**, 1146–1157 (2004).
- Bonneau, D. & Longy, M. Mutations of the human PTEN gene. *Hum. Mutat.* **16**, 109–122 (2000).
- Chi, Y. I. Homeodomain revisited: a lesson from disease-causing mutations. *Hum. Genet.* **116**, 433–444 (2005).
- Mortazavi, Y. *et al.* The spectrum of PIG-A gene mutations in aplastic anemia/paroxysmal nocturnal hemoglobinuria (AA/PNH): a high incidence of multiple mutations and evidence of a mutational hot spot. *Blood* **101**, 2833–2841 (2003).
- Iacobuzio-Donahue, C. A. *et al.* Missense mutations of MADH4: characterization of the mutational hot spot and functional consequences in human tumors. *Clin. Cancer Res.* **10**, 1597–1604 (2004).
- Galtier, N., Enard, D., Radondy, Y., Bazin, E. & Belkhir, K. Mutation hot spots in mammalian mitochondrial DNA. *Genome Res.* **16**, 215–222 (2005).
- Maki, H. Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annu. Rev. Genet.* **36**, 279–303 (2002).
- Arndt, P. F., Hwa, T. & Petrov, D. A. Substantial Regional Variation in Substitution Rates in the Human Genome: Importance of GC Content, Gene Density, and Telomere-Specific Effects. *J. Mol. Evol.* **60**, 748–763 (2005).
- Rogozin, I. B. & Pavlov, Y. I. Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat. Res.* **544**, 65–85 (2003).
- Silva, J. C. & Kondrashov, A. S. Patterns in spontaneous mutation revealed by human-baboon sequence comparison. *Trends Genet.* **18**, 544–547 (2002).
- Crow, J. F. The high spontaneous mutation rate: is it a health risk? *Proc. Natl Acad. Sci. USA* **94**, 8380–8386 (1997).
- Ellegren, H., Smith, N. G. & Webster, M. T. Mutation rate variation in the mammalian genome. *Curr. Opin. Genet. Dev.* **13**, 562–568 (2003).
- Denver, D. R., Morris, K., Lynch, M. & Thomas, W. K. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**, 679–682 (2004).
- Longman-Jacobsen, N., Williamson, J. F., Dawkins, R. L. & Gaudieri, S. In polymorphic genomic regions indels cluster with nucleotide polymorphism: Quantum Genomics. *Gene* **312**, 257–261 (2003).
- Petrov, D. A. DNA loss and evolution of genome size in *Drosophila*. *Genetica* **115**, 81–91 (2002).
- Hardison, R. C. *et al.* Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res.* **13**, 13–26 (2003).
- Kondrashov, A. S. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* **21**, 12–27 (2003).
- Dawkins, R. *et al.* Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol. Rev.* **167**, 275–304 (1999).
- The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
- Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- Strickberger, M. S. *Evolution* (Jones and Bartlett Publishers, 2000).
- Lercher, M. J. & Hurst, L. D. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* **18**, 337–340 (2002).
- Check, E. Human genome: patchwork people. *Nature* **437**, 1084–1086 (2005).
- Conrad, D. F., Andrews, T. D., Carter, N. P., Hurles, M. E. & Pritchard, J. K. A high-resolution survey of deletion polymorphism in the human genome. *Nature Genet.* **38**, 75–81 (2006).
- Schwartz, S. *et al.* Human-mouse alignments with BLASTZ. *Genome Res.* **13**, 103–107 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Kondrashov (a reviewer) for suggesting the 3-species test and the possibility of a signature for indel-caused substitutions; M. Long and X. Gao for comments; and T. Petes for informing us about genome-wide measures of recombination in yeast. This study was supported by NSFC (30470924, 30470122 & 30570987) and Pre-program for NBRPC (2005CCA02100) to D.T. or J.-Q.C. and by 111 Project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.T. (dtian@nju.edu.cn) or J.-Q.C. (chenjq@nju.edu.cn).

METHODS

Sequences and alignments. The alignments between human and chimpanzee, human and rhesus macaque, mouse and rat, and fruitfly were downloaded from UCSC (<http://hgdownload.cse.ucsc.edu>). Two rice and three yeast sequences were obtained from GRAMENE (<http://www.gramene.org>), BGI (<http://rice.genomics.org.cn>), SGD (<http://www.yeastgenome.org>), MIT (<http://www.broad.mit.edu>) or GTC (<http://sequence-www.stanford.edu>). The sequences for other species were from UCSC or GenBank. The rice and yeast sequences were aligned by BlastZ²⁵ (see the aligning flowchart in Supplementary Fig. 8). The BlastZ scoring matrix was the one UCSC used for their pairwise alignments of human and chimpanzee. The coding, non-coding, repeat and intron sequences are based on ENSEMBL and GRAMENE annotations.

To evaluate possible alignment artefacts, 11 Mb rice sequences and 11.4 Mb yeast alignments were aligned again manually and by ClustalW²⁶, respectively (Supplementary Figs 2 and 3), and the results compared. Alignments in areas with or without paralogous sequences and with or without transposons were analysed separately (Supplementary Fig. 4).

These alignments contain indels that are <101 bp (or <301 bp between rice lines and among yeast strains). When there is an indel >100 (or >300) bp or an ambiguous nucleotide, the aligned sequence was sectioned into two subalignments. Therefore, indel size ranges from 1 to 100 (or 300) bp and the aligned sequences contained no ambiguous nucleotides. To obtain a longer alignment, we removed all alignments that were <10 kb.

Analysis of nucleotide divergence. Each of the sequence alignments was dissected into small, non-overlapping windows, each having a specific distance to the nearest indel (d_1). An interval is defined as a region between two indels, the length of which reflects the reciprocal of the indel density (d_2). Only the intact intervals were used for analysis.

The windows in an interval are named and ordered as L0 (the first 50 bp or 50–100 when $d_2 = 100$ –199 bp), L1 (100 bp), ..., L n (and/or R n , 100–199 bp), ..., R1 (100 bp) and R0 (the last 50 bp), respectively (illustration in Supplementary Fig. 1a). An interval 11–99 bp long is assigned to W0' (intervals shorter than 11 bp are excluded from analysis). By definition, L1 or R1 is 0.1 kb closer than L2 or R2 to an indel, respectively, and L n (and/or R n) is located at the centre of an interval. The total length (d_2) of the contiguous, indel-free window is equal to L1 + ... L n + R n + ... + R1, in which the n varies among intervals. The nucleotide divergence (D) for each window is corrected by the Jukes–Cantor method²⁷.

Three-(or four)-lineage comparisons. We used an outgroup sequence from a closely related species (described in Supplementary Tables 2 and 3) to polarize each indel mutational difference between two (or three) ingroup lineages. This allowed us to identify the lineage in which indel events occurred. Following the same logic, we determined the number of nucleotide substitutions at fixed intervals from the indel in the lineage with the indel (N_i or D_i for divergence) and without the indel (N_{ni} or D_{ni}). High-quality three-(or four)-sequence alignments were obtained by first using BlastZ to find orthologous sequences, and then aligning them with ClustalW. To assure a conservative estimate, indels were discarded if the interval was <100 bp; only simple indels were included (excluded were complex indels, loci with >1 indels among three sequences or with ≥ 4 nucleotide repeats (the slippage-like indels)). The ratios $r = N_i/N_{ni}$ or $R = D_i/D_{ni}$ provide an estimate of the relative substitution rate difference attributable to the indel. In the absence of any association between an indel and the nucleotide substitution rate, the expected number of mutations on indel-containing and non-indel-containing branches is expected to be the same; with indel-associated mutation, we expect $N_i > N_{ni}$ and $D_i > D_{ni}$.

We used yeast strains S288C, RM11 and YJM789 to estimate the ratio r , and made the assumption in our model that these strains are representative of a population sample. Even though they were collected from different environments, the three strains are equally distantly diverged from one another, and equally share derived differences in a mosaic pattern across the genome, indicating extensive genetic exchange²⁸, suggesting that our assumption is reasonable.

26. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
27. Jukes, T. H. & Cantor, C. R. in *Mammalian Protein Metabolism* (ed. Munro, H. N.) 21–132 (Academic, 1969).
28. Ruderfer, D. M., Pratt, S. C., Seidel, H. S. & Kruglyak, L. Population genomic analysis of outcrossing and recombination in yeast. *Nature Genet.* **38**, 1077–1081 (2006).

Molecular architecture of native HIV-1 gp120 trimers

Jun Liu^{1*}, Alberto Bartesaghi^{1*}, Mario J. Borgnia^{1*}, Guillermo Sapiro² & Sriram Subramaniam¹

The envelope glycoproteins (Env) of human and simian immunodeficiency viruses (HIV and SIV, respectively) mediate virus binding to the cell surface receptor CD4 on target cells to initiate infection¹. Env is a heterodimer of a transmembrane glycoprotein (gp41) and a surface glycoprotein (gp120), and forms trimers on the surface of the viral membrane. Using cryo-electron tomography combined with three-dimensional image classification and averaging, we report the three-dimensional structures of trimeric Env displayed on native HIV-1 in the unliganded state, in complex with the broadly neutralizing antibody b12 and in a ternary complex with CD4 and the 17b antibody. By fitting the known crystal structures^{2,3} of the monomeric gp120 core in the b12- and CD4/17b-bound conformations into the density maps derived by electron tomography, we derive molecular models for the native HIV-1 gp120 trimer in unliganded and CD4-bound states. We demonstrate that CD4 binding results in a major reorganization of the Env trimer, causing an outward rotation and displacement of each gp120 monomer. This appears to be coupled with a rearrangement of the gp41 region along the central axis of the trimer, leading to closer contact between the viral and target cell membranes. Our findings elucidate the structure and conformational changes of trimeric HIV-1 gp120 relevant to antibody neutralization and attachment to target cells.

It is estimated that over 33 million individuals are at present infected with HIV (<http://www.unaids.org>). The development of an effective vaccine is therefore a challenge of fundamental medical interest. It has been widely recognized that a better understanding of the structure of trimeric Env in its various conformational states is likely to be an important element in the overall strategy for vaccine development⁴. Although X-ray crystallographic methods have led to atomic models for HIV-1 gp120 monomers complexed to antibodies in the presence and absence of CD4 (refs 2, 3, 5), determination of the structures of intact trimers on native viruses has nevertheless remained elusive. Theoretical models for the structure of the trimer that take into account constraints determined from biochemical and mutagenesis studies of monomeric gp120 (refs 6, 7) have been advanced, but the advent of electron tomographic methods⁸ provides a unique opportunity for direct experimental determination of the structure of the intact trimer on the virus in its native state. Here we report structural analysis of native HIV-1 Env using alignment and classification procedures that take into account the missing wedge that arises from the limited angular range used for data collection in electron tomography. Our approach takes advantage of complexes containing monomeric gp120 for which there are known X-ray structures, allowing us to derive models for trimeric gp120 in unliganded and CD4-bound states.

We first analysed tomograms obtained with viruses complexed with Fab fragments from the potent, broadly cross-reactive, neutralizing antibody b12, as an atomic model of the complex of a disulphide-bond stabilized version of monomeric gp120 core with the Fab

fragment of b12 is available³ (see Supplementary Methods and Supplementary Figs 1–3 for a detailed description of methods). The contributions of the Fab fragment to the experimentally derived density map can be easily spotted (Fig. 1). The X-ray coordinates for the gp120–Fab complex were docked as a rigid body into the map using automated fitting procedures (Supplementary Fig. 4), resulting in a description of the molecular structure of the b12-complexed trimer. The X-ray structure of monomeric gp120 in complex with b12–Fab includes only ~58% of the gp120 polypeptide sequence, and lacks most of the residues in the V1/V2 loops (residues 121–203), V3 loop (residues 300–328) and portions of the amino and carboxy termini (residues 1–82 and 493–511, respectively). Inspection of the extra densities in the density map that are not occupied by the coordinates reveal the likely locations of these regions, as well as the probable location of gp41 in the native trimer (Fig. 1). In particular,

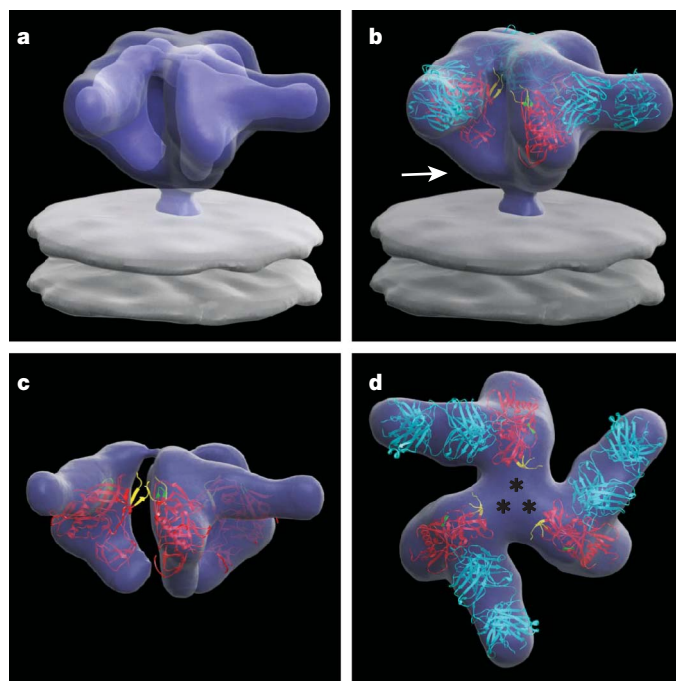


Figure 1 | Averaged 3D structure of the HIV-1 spike in complex with b12–Fab. **a**, Perspective view of the surface of the density map shown at two thresholds; one to include the entire spike (outer), and another to highlight the Fab and gp120 components (inner). **b–d**, Front (**b**, **c**) and top (**d**) views of the map fitted with X-ray coordinates of the complex of the Fab fragment of b12 (cyan) with gp120 (red, PDB ID, 2NY7); only gp120 coordinates are shown in **c**, which is at the inner threshold. Likely locations of the V1/V2 loop and gp41 regions are indicated by asterisks in **d** and the white arrow in **b**, respectively. The stumps of the V1/V2 and V3 loop regions are shown in yellow and green, respectively.

¹Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA. ²Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA.

*These authors contributed equally to this work.

the unassigned densities adjacent to the V1/V2 stem have a size consistent with that expected from the ~ 80 residues missing in the V1/V2 loop, implying that the three V1/V2 loop regions come together to form the apex of the mushroom-shaped Env trimer.

Analysis of the native, undecorated Env trimer (Fig. 2a–c) shows the general shape and arrangement of gp120 monomers in the native spike, which are comparable to those obtained for the Env–b12 complex (also see Supplementary Video). Overall, the spike has a height of ~ 120 Å, and a maximal width of ~ 150 Å, which tapers from ~ 80 Å at the base of the gp120 regions to ~ 35 Å at the junction with the membrane. The best fit of gp120 into the density map is shown in Fig. 2d, e with the conformation of gp120 derived in complex with b12 (identical to gp120 coordinates shown in Fig. 1), and from the conformation obtained in a complex with the X5 antibody⁵ (Fig. 2f; see Supplementary Fig. 4 for a detailed description of the fitting procedures used). The latter complex includes the V3 loop region (in the CD4-bound conformation, which is likely to be very different from the conformation in the unliganded state), but the gp120 constructs used to obtain both crystal structures lack significant portions of the V1/V2 loop and the N and C termini of gp120. The regions of the V1/V2 stem that are included in the structure display high temperature factors indicating potential flexibility², and also may not reflect their actual positions in the intact loop region because of the truncation. Nevertheless, inspection of the fits confirms and extends the general conclusion drawn from Fig. 1—that the V1/V2 and V3 regions on each monomer are near the apex of the trimer, and that the three gp41 components form a mushroom-shaped structure at the base of the gp120 trimer. The residues likely to be glycosylated, as well as the variable loops (V1–V4), are all generally located in solvent accessible regions, as suggested in a previously proposed theoretical model for trimer architecture⁶.

Understanding the nature of conformational changes in the Env trimer induced by CD4 binding⁹ is at the heart of defining the molecular mechanisms underlying HIV entry into cells. We therefore carried out electron tomographic analysis of HIV-1 complexed to

CD4 and the Fab fragment of the 17b antibody, both because the 17b antibody has been shown¹⁰ to stabilize and lock gp120 in the CD4-bound conformation, and because the crystallographic structure of the ternary complex of gp120 with CD4 and the 17b Fab fragment has been determined². Inspection of the averaged structure of the Env trimer bound to CD4 and 17b (Fig. 3a, b) shows a dramatic conformational change in comparison with that of the unliganded trimer (Fig. 2). As in the case of the previous maps, the constraints provided by the presence of CD4 and 17b ensure that there is a single, unambiguous fit of the coordinates into the density map (see Supplementary Fig. 4), and provide the basis to understand the conformational change induced by CD4 binding. There is clear additional density in the averaged map for the V1/V2 loop (~ 70 residues) that is missing in the crystal structure of the complex (Fig. 3a). Relative to gp120 in the unliganded trimer, each gp120 monomer in the CD4 complex displays a rotation of about 45° around an axis parallel to the central three-fold axis, an out of plane rotation of about 15° , and an upward displacement of the overall centre of mass by ~ 15 Å. There are also discernible changes in the gp41 region adjacent to the viral membrane, and a new feature is observed at the centre of the spike that is not present in the unliganded spike or in the complex with b12. The most likely interpretation of this feature is that it arises from rearrangements of gp41 that eventually lead to formation of the six-helix bundle structure, and towards fusion between viral and target cell membranes¹¹.

Comparison of the locations of the docked gp120 monomers in the free, b12-bound and CD4-bound states (Fig. 3c–e) provides insights into the overall quaternary structural changes that occur in the trimeric spike. The binding of b12 results in a partial opening of the spike, coupled with rotation of each monomer by $\sim 20^\circ$ – 25° around an axis perpendicular to the viral membrane (Fig. 3d). However, CD4 binding results in a rotation around this central axis in the same direction that is twice as large, in addition to an out-of-plane rotation (Fig. 3e), and slight vertical displacement of gp120. Thus, while the binding sites for CD4 and b12 are on roughly the same face of the

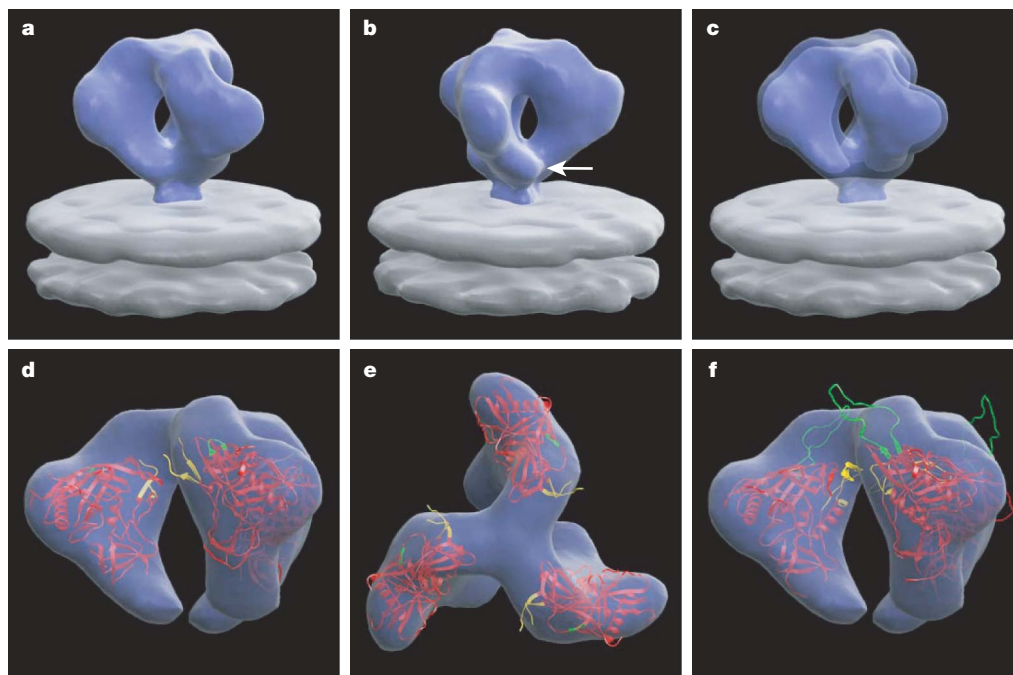


Figure 2 | Averaged 3D structure of the trimeric glycoprotein spike on native HIV-1. **a, b,** Perspective/front views of the surface of the density map; the white arrow in **b** points to the likely location of gp41 in the map. **c,** Same view as in **a** but shown using two thresholds to illustrate both the overall shape (outer), and the contribution of the gp120 containing regions (inner). **d, e,** Front and top views of the map with coordinates for the gp120 core

derived from the complex with b12 (PDB ID, 2NY7). **f,** Front view of the map fitted with coordinates for the gp120 core derived from the complex with X5 (PDB ID, 2B4C). The gp120 core is shown in red, and the regions of the V1/V2 loop and V3 loop included in the coordinates are shown in yellow and green, respectively.

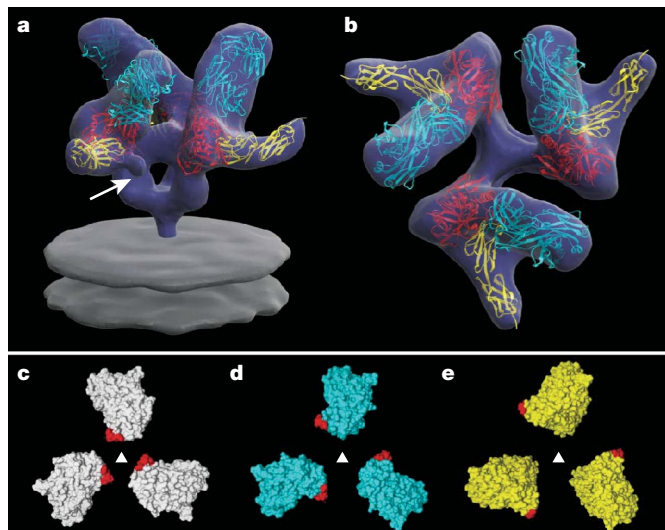


Figure 3 | 3D structure of the HIV-1 spike in complex with CD4 and 17b-Fab. **a, b**, Front and top views showing the X-ray coordinates of the ternary complex (PDB ID, 1GC1) of the gp120 core (red) with CD4 (yellow) and the Fab fragment of 17b (cyan) fitted into the map as a rigid body using automated fitting procedures²⁷. The arrow in **a** points to the likely location of the V1/V2 loop region, which was partially deleted in the construct used to obtain crystals of the ternary complex². **c–e**, Top views showing superposition of the X-ray coordinates for the gp120 trimer derived from the maps for the unliganded (white), b12-bound (cyan) and CD4/17b-bound (yellow) states of the trimeric spike, with locations of the V1/V2 stem regions indicated in red.

gp120 monomer, they result in very different outcomes for the conformation of the Env trimer.

The observed fit of the gp120 core regions from CD4-liganded gp120 complexes (with X5 as well as 17b antibodies) into the corresponding regions in the density map of the native trimer has important implications for the nature and extent of structural differences in gp120 that occur upon CD4 binding. So far, there are no reports of the atomic structures of either monomeric or trimeric HIV-1 envelope glycoproteins determined using X-ray crystallography. Chen *et al.*⁷ reported a structure for a truncated, unliganded SIV gp120 monomer in which the conformation of gp120 is different from that of the HIV-1 gp120 monomer seen in either the CD4-liganded or b12-bound states⁷. Our findings show that the conformation of the HIV-1 gp120 monomer observed in the b12 and CD4-liganded states can be docked into density maps of the unliganded HIV-1 spike, whereas the conformation reported for the SIV gp120 monomer does not represent a good fit (Supplementary Figs 5 and 6). Further, the trimeric gp120 arrangement we have derived has the V1/V2 loop regions at the apex of the trimer, in contrast to the arrangement suggested by Chen *et al.*⁷, in which the V1/V2 loop regions are close to the base of the gp120 trimer. Probable explanations for these differences include the possibility that the three-dimensional (3D) crystals used to determine the structure of the truncated SIV gp120 core captured a conformation of gp120 that is different from the physiologically relevant conformation in the native trimer, or that there are fundamental differences in conformation between monomeric SIV and HIV-1 gp120. We note that the density map we have obtained for the HIV-1 spike has resemblance in the stalk region to the map reported by Zanetti *et al.*¹² for the SIV spike, and the overall features of this map are comparable to a low-resolution version of the map we show in Fig. 2 (see Supplementary Fig. 3 for progressive improvement in map resolution with iterative refinement). However, our results are at variance with the conclusion of Zhu *et al.*¹³ that the membrane-proximal region of gp41 is splayed out into three distinct 'legs' separated by ~80 Å from each other at the point of contact with the membrane (Supplementary Fig. 7).

The dramatic 'opening' of the trimer induced by CD4 has profound consequences for the disposition of the various key regions of the spike relative to the viral membrane and the target cell (Fig. 4). Previous measurements of the energetics of the CD4 binding have suggested the existence of a large entropic contribution (conformational 'fixation' of gp120) that results from CD4 binding^{10,14}. At the present resolution of our maps (~19, 22 and 23 Å respectively for the maps shown in Figs 1–3, as determined by the 0.5 FSC criterion, see Supplementary Fig. 1), we cannot yet directly determine the nature of conformational changes in the monomer, but our analysis implies that in addition to any such changes in the monomer, there are large, additional contributions from quaternary changes in the structure of the trimer. The lever-like opening of the trimer upon CD4 binding makes way for exposure of the central gp41 stalk. The V3 loop region is released from the lateral edge of the apex of the spike to directly point towards the target cell, while the V1/V2 regions as well as the CD4 binding sites move away from the centre of the spike (Fig. 4b).

The determination of molecular models for trimeric gp120 in the free and CD4-complexed states could represent a useful starting point for the development of rationally designed vaccines to counter the AIDS epidemic that take into account the trimeric structure of the spike. In the native state, the trimer is held together by strong contacts at the gp41 base and at the apex, which appear to have significant contributions from the V1/V2 loop regions, and are adjacent to the V3 loop region and chemokine receptor binding sites¹⁵. The locations of these regions in the unliganded spike at the critical interface between virus and the target cell are consistent with observations documenting the altered susceptibility of variants with deletions in the V1 and V2 loop regions to neutralization^{16–18} as well as the identification of antibodies to unique quaternary epitopes on Env¹⁹. There appears to be relatively little contact between most other regions of the neighbouring gp120 monomers, resulting in a spike architecture that is held together somewhat tenuously at the top and bottom, poised to be sprung open upon CD4 binding. The CD4 binding site is recessed by about 20 Å from the top of the spike, with the V1/V2 regions and associated carbohydrate moieties forming a sheath at the top. The recessed site implies that cell surface CD4 must delve into the spike to achieve binding. The outward movement of gp120 results in a steep change in the orientation of the two outermost domains (D1D2) of CD4 (Fig. 4e), implying that this motion must draw the virus closer to the target cell membrane by virtue of the flexibility between the D1D2 and D3D4 domains of membrane-anchored CD4 (Fig. 4f). Indeed, recent cryo-electron tomographic studies of the complex of native viral gp120 with D1D2-IgP suggest that the highly potent neutralizing activity of D1D2-IgP probably arises from its flexible, polyvalent nature, which allows its binding to multiple spikes on the same virus and across neighbouring viruses²⁰. A prediction of our model is that the hinge region between the second and third extracellular domains of CD4 is critical for viral entry, potentially explaining observations that antibodies directed against epitopes close to this hinge region block fusion and HIV infection²¹, and that binding to gp120 induces a dramatic bend in CD4 at this hinge region²².

The nature and extent of the CD4-induced structural change that we have identified here provides a structural foundation to interpret and refine the plethora of biochemical studies on Env and to better understand mechanisms that are important for virus neutralization and entry^{4,23,24}. We propose from our results that CD4 binding draws the spike closer to the target cell membrane as a result of the hinge-like motion of the D1D2 domains resulting from the change in trimer structure. It is clear from this sequence of events that the exposure of the V3 loop and other antigenic determinants important for viral attachment occur in the protected milieu of the interface formed between the viral and target cell membrane, providing a mechanism for seclusion of these epitopes from antibodies whose binding could potentially neutralize HIV-1.

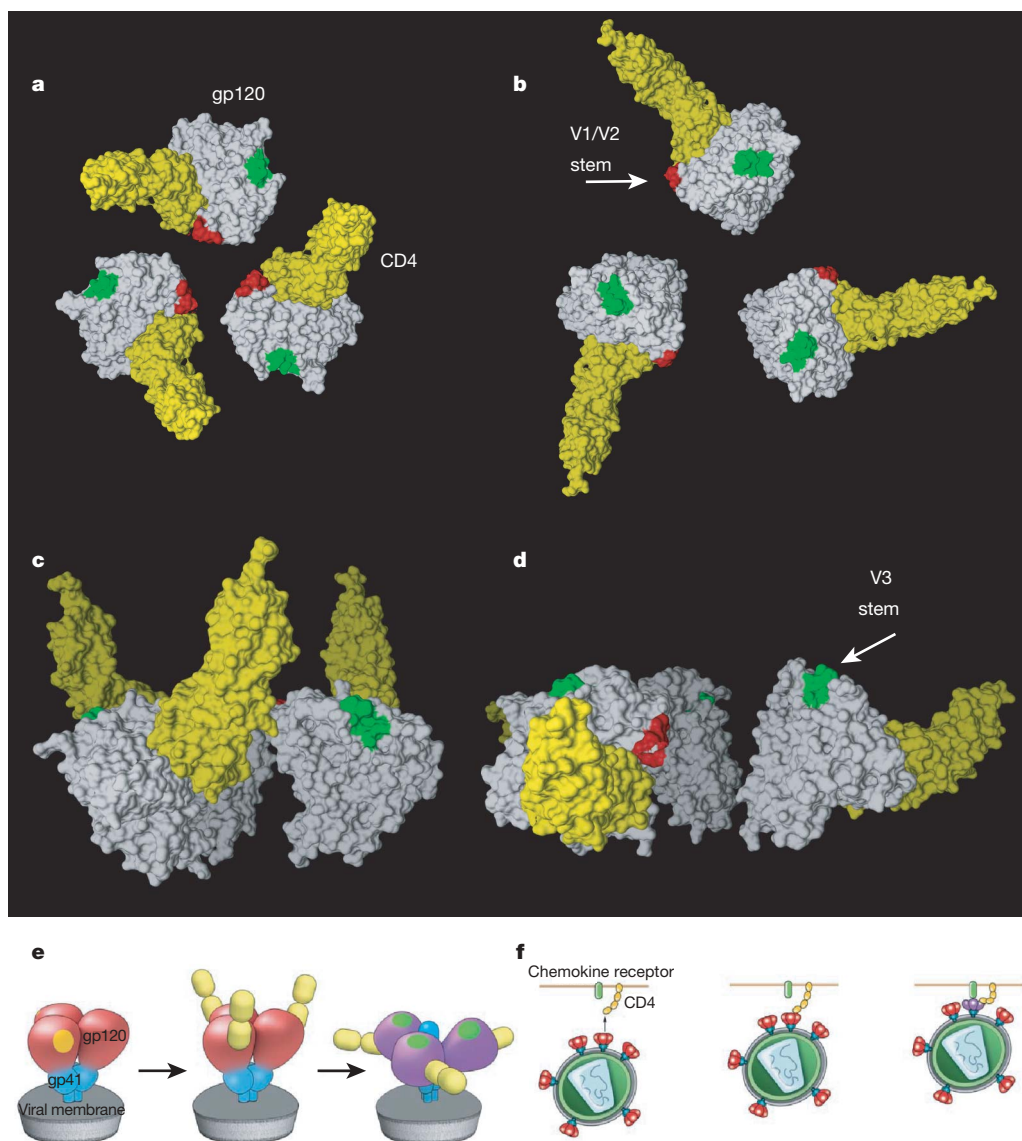


Figure 4 | Description of the conformational change in the gp120 trimer induced by CD4 binding. **a–d**, Model for the conformational change from the unliganded (**a**, **c**) to the CD4-bound state (**b**, **d**) shown as top (**a**, **b**) and front (**c**, **d**) views. The gp120 core, CD4, V1/V2 and V3 stems are shown in white, yellow, red and green colours, respectively. **e**, Schematic description of the gp41 (blue) and gp120 (red/purple) regions of the trimeric spike and the conformational changes that occur upon CD4 binding. The yellow patch

near the apex marks the location of the CD4 binding site in the unliganded spike and the green patch at the apex marks the location of the V3 loop region in the spike after CD4 binding. **f**, Schematic view of the consequence of the CD4-induced conformational changes for viral attachment to the target cell and interaction with chemokine receptors (green at top). Colours in **f** have same meaning as in **e**.

Given the limited accessibility of key antigenic sites on gp120, the observed rearrangement of the gp120 trimer upon binding of the broadly neutralizing antibody b12 is a surprising and potentially important finding. The outward displacement of gp120 in the b12-bound state appears to be along the same general trajectory observed upon CD4 binding (Fig. 3d, e), but appears to lock gp120 and trimeric Env in a state that prevents further conformational changes that could lead to exposure of the V3 loop or rearrangement of gp41. It is possible that the observed outward movement of gp120 is driven by conformational changes at the gp120–b12 interface³ that occur after initial contact with b12 in order to accommodate the steric consequences of the b12 binding in the context of an intact trimer. Knowledge of the trimeric structure of the HIV-1 spike in strains with differing levels of neutralization sensitivity, and at various stages of activation that ultimately culminate in formation of the viral entry claw²⁵, will be important to further understand the range of variation in spike structure. Approaches for intervention that target regions on

Env that are critical for the conformational change could provide a new addition to the arsenal of strategies to combat HIV/AIDS.

METHODS SUMMARY

Purified viruses in the presence or absence of added reagents (antibodies, CD4) were deposited on home-made holey carbon grids and plunge-frozen in liquid ethane maintained at about -180° to prepare vitrified specimens for cryo-electron tomography. Specimens were imaged in a Polara transmission electron microscope equipped with an energy filter, with the specimen maintained at liquid nitrogen temperatures. Typically, a series of 141 low dose images of each frozen hydrated virus was recorded at 1° tilt intervals in the range of $\pm 70^{\circ}$. The tilt series were aligned and back-projected to reconstruct 3D volumes (tomograms) of individual viruses. Viral spikes protruding from the membrane surface were readily identified in the tomograms (Supplementary Fig. 1), and were extracted for further processing. In total, we extracted 4,741 spikes from 382 virions of HIV-1 strain BaL, 4,323 spikes from 306 virions of HIV-1 BaL complexed with b12-Fab, and 4,900 spikes from 503 virions of HIV-1 BaL complexed with CD4 and 17b-Fab. Alignment, classification and 3D averaging of the

extracted subvolumes for determination of the 3D structure were carried out based on the framework described in ref. 26 (see Supplementary Fig. 1a, b for representative examples of a tomographic slice and segmented virus, respectively, and Supplementary Figs 2 and 3 for an illustration of classes and class averages at early and late stages of refinement). No external references were used for alignment and classification, and the presence of the missing wedge of information in each volume was taken into account for 3D alignment. Iterative 3D classification and alignment runs were executed starting with the raw images until no further changes were observed in the final density maps. Fitting of coordinates into the density maps (Supplementary Figs 4–6) was carried out using automated procedures implemented in the software package Chimera²⁷.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 March; accepted 10 June 2008.

Published online 30 July 2008.

- Dalgleish, A. G. *et al.* The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**, 763–767 (1984).
- Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
- Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732–737 (2007).
- Burton, D. R. *et al.* HIV vaccine design and the neutralizing antibody problem. *Nature Immunol.* **5**, 233–236 (2004).
- Huang, C. C. *et al.* Structure of a V3-containing HIV-1 gp120 core. *Science* **310**, 1025–1028 (2005).
- Kwong, P. D. *et al.* Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus. *J. Virol.* **74**, 1961–1972 (2000).
- Chen, B. *et al.* Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **433**, 834–841 (2005).
- Subramaniam, S. Bridging the imaging gap: Visualizing subcellular architecture with electron tomography. *Curr. Opin. Microbiol.* **8**, 316–322 (2005).
- Sattentau, Q. J. & Moore, J. P. Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J. Exp. Med.* **174**, 407–415 (1991).
- Kwong, P. D. *et al.* HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* **420**, 678–682 (2002).
- Chan, D. C. & Kim, P. S. HIV entry and its inhibition. *Cell* **93**, 681–684 (1998).
- Zanetti, G. *et al.* Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. *PLoS Pathog.* **2**, e83 (2006).
- Zhu, P. *et al.* Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* **441**, 847–852 (2006).
- Myszka, D. G. *et al.* Energetics of the HIV gp120–CD4 binding reaction. *Proc. Natl Acad. Sci. USA* **97**, 9026–9031 (2000).
- Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393**, 705–711 (1998).
- Cao, J. *et al.* Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J. Virol.* **71**, 9808–9812 (1997).
- Johnson, W. E. *et al.* A replication-competent, neutralization-sensitive variant of simian immunodeficiency virus lacking 100 amino acids of envelope. *J. Virol.* **76**, 2075–2086 (2002).
- Saunders, C. J. *et al.* The V1, V2, and V3 regions of the human immunodeficiency virus type 1 envelope differentially affect the viral phenotype in an isolate-dependent manner. *J. Virol.* **79**, 9069–9080 (2005).
- Gorny, M. K. *et al.* Identification of a new quaternary neutralizing epitope on human immunodeficiency virus type 1 virus particles. *J. Virol.* **79**, 5232–5237 (2005).
- Bennett, A. E. *et al.* Cryo electron tomographic analysis of an HIV neutralizing protein and its complex with native viral gp120. *J. Biol. Chem.* **282**, 27754–27759 (2007).
- Healey, D. *et al.* Novel anti-CD4 monoclonal antibodies separate human immunodeficiency virus infection and fusion of CD4+ cells from virus binding. *J. Exp. Med.* **172**, 1233–1242 (1990).
- Ashish *et al.* Conformational rearrangement within the soluble domains of the CD4 receptor is ligand-specific. *J. Biol. Chem.* **283**, 2761–2772 (2008).
- Montefiori, D. *et al.* Antibody-based HIV-1 vaccines: Recent developments and future directions. *PLoS Med.* **4**, e348 (2007).
- Zwick, M. B. & Burton, D. R. HIV-1 neutralization: Mechanisms and relevance to vaccine design. *Curr. HIV Res.* **5**, 608–624 (2007).
- Sougrat, R. *et al.* Electron tomography of the contact between T cells and SIV/HIV-1: implications for viral entry. *PLoS Pathog.* **3**, e63 (2007).
- Bartesaghi, A. *et al.* Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.* **162**, 436–450 (2008).
- Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
- Rossio, J. L. *et al.* Inactivation of human immunodeficiency virus type 1 infectivity with preservation of conformational and functional integrity of virion surface proteins. *J. Virol.* **72**, 7992–8001 (1998).
- Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
- Winkler, H. 3D reconstruction and processing of volumetric data in cryo-electron tomography. *J. Struct. Biol.* **157**, 126–137 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Bess and J. Lifson for providing purified, AT-2 treated HIV-1 BaL virus; P. Kwong for providing purified preparations of Fab fragments of the neutralizing antibodies b12 and 17b; S. Fellini and colleagues for assistance with use of the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Maryland (<http://biowulf.nih.gov>); E. Tyler for assistance with figures; T. M. Chou for assistance with microscope maintenance; and A. Bennett, J. Milne and T. White for comments on the manuscript. This work was supported by funds from the intramural programme of the National Cancer Institute (to S.S.), and from the National Science Foundation and the Department of Defense (to G.S.).

Author Information Data are available as accession numbers EMD-5018 to EMD-5023 in the EM Data Bank, with associated PDB entries 3DNL, 3DNN and 3DNO in the Protein Data Bank. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.S. (ss1@nih.gov).

METHODS

Reagents. Samples of HIV-1 strain BaL (estimated concentration $\sim 10^{11}$ virions ml^{-1}), purified by sucrose gradient centrifugation and inactivated by treatment with Aldrichthiol-2 (AT-2), were contributed by J. Bess and J. Lifson. AT-2 treated viruses are capable of supporting viral entry at levels comparable to untreated viruses, and have a similar antigenic profile to untreated viruses²⁸. Purified soluble CD4 (sCD4; 1–183 containing fragment) was obtained from the NIH AIDS reagent program, while purified Fab fragments from b12 and 17b antibodies were provided by P. Kwong.

Specimen preparation. Purified viral suspensions were pre-incubated at 4 °C for 15–30 min in buffer alone, or in the presence of (1) the Fab fragment of the b12 antibody, and (2) sCD4 and the Fab fragment of 17b. All ligands were added at a concentration corresponding to an estimated fivefold molar ratio with Env trimers. Samples were then mixed with 5-nm colloidal gold (used as fiducial markers in initial image alignment) and deposited on home-made, carbon-coated holey carbon grids. Excess liquid was blotted with filter paper from both sides of the grid to form a thin layer of buffer which was then rapidly frozen by plunging the grid in a liquid/solid ethane slush (about -180 °C). This procedure, which results in the embedding of the viruses in a ~ 150 nm layer of amorphous ice spanning holes in the carbon layer, was carried out using a Vitrobot rapid freezing device (FEI).

Cryo-electron tomography. Frozen virus specimens were imaged at liquid nitrogen temperatures using a Polara field emission gun electron microscope (FEI) equipped with a $2\text{ k} \times 2\text{ k}$ CCD placed at the end of GIF 2002 energy filter (Gatan), operated in the zero-energy-loss mode with a slit width of 20 eV. The microscope was operated at 200 kV and a magnification of $\times 34,000$, resulting in an effective pixel size of 4.1 Å. Tilt series were collected in automatic batch mode. Low dose single-axis tilt series were collected from each virus specimen at nominal underfocus settings of -2 μm . Since the defocus was determined in regions on the carbon film, and ~ 3 – 4 μm away from the imaged viruses located in the vitreous ice, we estimate that the actual defocus of the collected data ranges from about -1.5 to -2.5 μm . Under these conditions, the first zero of the contrast transfer function (CTF) is at ~ 22 Å. The angular range of the tilt series was from -70° to $+70^\circ$, typically at tilt increments of 1° , and at a cumulative dose of ~ 80 electrons \AA^{-2} . Tilt series were initially aligned with gold markers using Inspect3D (FEI), and reconstructed after further refinement using weighted back-projection as implemented in the software packages IMOD²⁹ and Protomo³⁰. Visualization of tomograms was carried out using software tools implemented in the program Amira (TGS Inc) and UCSF Chimera.

Classification and 3D averaging. Volumes corresponding to reconstructions of individual viruses were extracted from the tomograms. Locations of surface spikes on each virion were identified by manual inspection, and the corresponding subvolumes ($128 \times 128 \times 128$ voxels) were computationally extracted. Densities on the surface much smaller than the expected size of the trimer (~ 120 Å high and ~ 150 Å wide) were not selected. The approximate local orientations of the long axis of the spike were determined by fitting an ellipsoidal surface to the picked spike positions, with the surface normals at each measured point providing initial estimates for two of the three Euler angles. The remaining in-plane rotation was initially randomized to eliminate possible bias in initial

alignment. After application of the Euler angles, sub-volumes were translationally aligned to their cylindrically averaged global average so that all sub-volumes in a given data set shared the same centre of mass. Alignment and classification of the spike volumes was carried out using the framework described in ref. 26. Early stages of refinement clearly showed the inherent three-fold symmetry in the spike structure (see Supplementary Figs 2 and 3), and once this was visually ascertained, three-fold symmetry was imposed for subsequent rounds of refinement. At each round, the classes that showed the most clearly delineated features in all regions of the spike (typically $\sim 50\%$ – 60%) were selected to be used as references for the next round (for example, the top five classes in Supplementary Fig. 2 that show well-resolved densities for both the spike and the Fab fragment). All images were retained until the final iteration to allow for movement between classes. The final maps shown in Figs 1–3 were obtained after ~ 10 – 12 refinement rounds and include contributions from $\sim 50\%$ of the sub-volumes in each case. Fourier shell correlation coefficients were estimated by comparing the correlation between two randomly divided halves of the aligned images used to generate the final map. No temperature factor sharpening was applied to the final maps, and no correction for CTF was applied. Note that the resolution obtained is roughly the same as the resolution corresponding to the first zero of the CTF under the conditions used for data collection.

Fitting of coordinates into map. Surface renderings of all the maps, and automated fitting of atomic coordinates, was carried out in the environment of the software package Chimera²⁷. The gp120 complexes with b12 and CD4/17b were docked as rigid bodies, that is, the coordinates used are identical to the structures of the complex derived by X-ray crystallography. The coordinates were filtered to 20 Å resolution before carrying out the fits to match the resolution of the experimental density maps; however, the use of coordinates without filtering, or with filtering to intermediate resolutions, had little effect on the overall fit. The fits shown in Fig. 2 were done directly using 2NY7 (Fig. 2d, e) and 2B4C (Fig. 2f) coordinates, after verifying that the results were similar to those obtained using 1GC1 coordinates (used to derive the molecular model presented in Fig. 4). To obtain the superposition shown in Fig. 3c–e, the three density maps were first aligned to each other to establish a single frame of reference for the three sets of fitted coordinates. The coordinates of the gp120 component of 1GC1 (CD4/17b complex) in Fig. 3e were then directly fitted into the map of the unliganded spike to obtain Fig. 3c, and aligned to gp120 coordinates 2NY7 (b12 complex) to obtain Fig. 3d. The hypervariable loop regions V1, V2 and V3 were not considered for arriving at the final fits, and inclusion or exclusion of the residues present in the V3 loop region (2B4C coordinates) or the stump of the V1/V2 loop region (1GC1 coordinates) did not alter the results. The best fit of HIV-1 gp120 coordinates into the map is unambiguous (Supplementary Fig. 4). This is further supported by the geometric fitting exercises shown in Supplementary Fig. 5 that illustrate comparative analysis of fits to different sets of HIV-1 and SIV gp120 coordinates, and in Supplementary Fig. 6 showing comparative analysis of fit quality visualized over a wide range of thresholds. To arrive at the molecular model shown in Fig. 4, the coordinates of gp120 in the CD4/17b complex were fitted to the map of the unliganded spike (as in Fig. 3c) and the coordinates for CD4 were then placed in the same relative orientation to gp120 as observed in the X-ray structure of the gp120/CD4/17b complex to derive Fig. 4a and c.

LETTERS

Neurogenin 2 controls cortical neuron migration through regulation of *Rnd2*

Julian Ik-Tsen Heng¹, Laurent Nguyen¹, Diogo S. Castro¹, Céline Zimmer¹, Hendrik Wildner¹, Olivier Armant¹, Dorota Skowronska-Krawczyk², Francesco Bedogni³, Jean-Marc Matter², Robert Hevner³ & François Guillemot¹

Motility is a universal property of newly generated neurons. How cell migration is coordinately regulated with other aspects of neuron production is not well understood. Here we show that the proneural protein neurogenin 2 (*Neurog2*), which controls neurogenesis in the embryonic cerebral cortex^{1,2}, directly induces the expression of the small GTP-binding protein *Rnd2* (ref. 3) in newly generated mouse cortical neurons before they initiate migration. *Rnd2* silencing leads to a defect in radial migration of cortical neurons similar to that observed when the *Neurog2* gene is deleted. Remarkably, restoring *Rnd2* expression in *Neurog2*-mutant neurons is sufficient to rescue their ability to migrate. Our results identify *Rnd2* as a novel essential regulator of neuronal migration in the cerebral cortex and demonstrate that *Rnd2* is a major effector of *Neurog2* function in the promotion of migration. Thus, a proneural protein controls the complex cellular behaviour of cell migration through a remarkably direct pathway involving the transcriptional activation of a small GTP-binding protein.

Neurons migrate extensively after their birth to reach their permanent location in the nervous system^{4,5}. Several neurological diseases are caused by defects in neuronal migration, underlining the importance of this process for normal brain function^{6,7}. Proneural transcription factors, which coordinate the developmental program that drives the differentiation of neural stem cells into neurons^{8,9}, have also been shown to promote the radial migration of neurons in the embryonic cerebral cortex^{10–12}. However, the mechanisms underlying the migration-promoting activity of proneural proteins have not been elucidated and it is unclear how many downstream genes are involved. Several genes with important roles in cell migration, including *RhoA*, *doublecortin* (*Dcx*) and *p35*, have been proposed to mediate *Neurog2* function in cortical neuron migration^{10,11}. However, expression of these genes is only mildly affected in the cortex of *Neurog2* single-mutant or *Neurog1*; *Neurog2* double-mutant embryos¹³ (Supplementary Fig. 1a–i), suggesting that they have a minor contribution to the migration-promoting activity of *Neurog2*. We therefore searched for new targets of *Neurog2* that may promote neuronal migration in the cerebral cortex.

Our strategy to identify genes regulated by *Neurog2* in the embryonic cortex is schematized in Supplementary Fig. 1j and has been reported elsewhere¹⁴. Briefly, we performed an expression microarray analysis of the dorsal telencephalon in *Neurog2*- and *Neurog1*; *Neurog2*-mutant embryos and in wild-type embryos overexpressing *Neurog2*. Genes that showed reciprocal changes to their expression in these loss-of-function and gain-of-function experiments were then selected and their functional annotation in Gene Ontology (<http://www.geneontology.org>) was examined. From this screen, the gene *Rnd2*, which encodes a small GTP-binding protein^{15,16}, was selected as a candidate target of *Neurog2* which could potentially regulate the

cytoskeleton³ (Supplementary Fig. 1k). *Rnd2* has restricted expression in the embryonic cerebral cortex throughout development, with transcripts detected in scattered cells within the ventricular zone, where radial glial progenitor cells and newborn neurons are located, and in intermediate progenitors and migrating neurons populating the subventricular zone and intermediate zone; in the cortical plate, where neurons settle and differentiate, its expression is sharply downregulated (Fig. 1a–c, g). *Rnd2* expression was significantly downregulated in the cortex of *Neurog2*-mutant embryos and absent in *Neurog1*; *Neurog2*-mutant embryos (Fig. 1d–f). Its expression was induced when *Neurog2* was overexpressed in the dorsal telencephalon (Supplementary Fig. 2). *Rnd2* is thus transiently expressed by migrating cortical neurons and their immediate precursors, and this expression is controlled by *Neurog2*.

To study the function of *Rnd2* in cortical development, we electroporated small interference RNAs (siRNAs) together with enhanced green fluorescent protein (EGFP) expression construct into the cerebral cortex of day 14.5 mouse embryos *ex vivo* followed by organotypic slice culture for 4 days¹⁰ or *in utero*¹². A mutant form of *Rnd2* that cannot bind GTP does not have dominant-negative activity and therefore cannot be used to study *Rnd2* function (Supplementary Fig. 3 (refs 3, 17)). siRNA electroporation led to effective and specific *Rnd2* knockdown (Supplementary Figs 4 and 5a, b), and resulted in a striking defect in the radial migration of cortical projection neurons (Fig. 2a–c). Although 36% cells electroporated with a control siRNA were found in the cortical plate after 4 days, only 14% of cells electroporated with *Rnd2* siRNAs migrated to the same extent ($n = 6$). All stages of radial migration through the cortex were affected by *Rnd2* knockdown (that is, from the ventricular zone and subventricular zone to the intermediate zone, from the intermediate zone to the cortical plate, and from the lower cortical plate to the upper cortical plate; Supplementary Fig. 6). Electroporation of siRNAs at embryonic day (E)12.5, resulting in the suppression of *Rnd2* expression in neurons born earlier in cortical development, also affected migration (Supplementary Fig. 7). The proliferation of cortical progenitors, their specification to a cortical neuron identity and the organization of the radial glia scaffold, along with which cortical projection neurons migrate, all remained unperturbed, indicating that *Rnd2* silencing interfered specifically and cell autonomously with neuronal migration (Supplementary Figs 8 and 9).

In addition to migration defects, *Rnd2*-silenced neurons also displayed abnormal morphologies. During cortical development, newborn projection neurons normally adopt a transient multipolar morphology while migrating through the subventricular and intermediate zones^{7,18}. siRNA-mediated knockdown of endogenous *Rnd2* increased the fraction of intermediate zone neurons with a multipolar

¹Division of Molecular Neurobiology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK. ²Department of Biochemistry, Sciences II, University of Geneva, CH-1211 Geneva 4, Switzerland. ³Department of Pathology, University of Washington School of Medicine, Harborview Medical Center, Seattle, Washington 98104, USA.

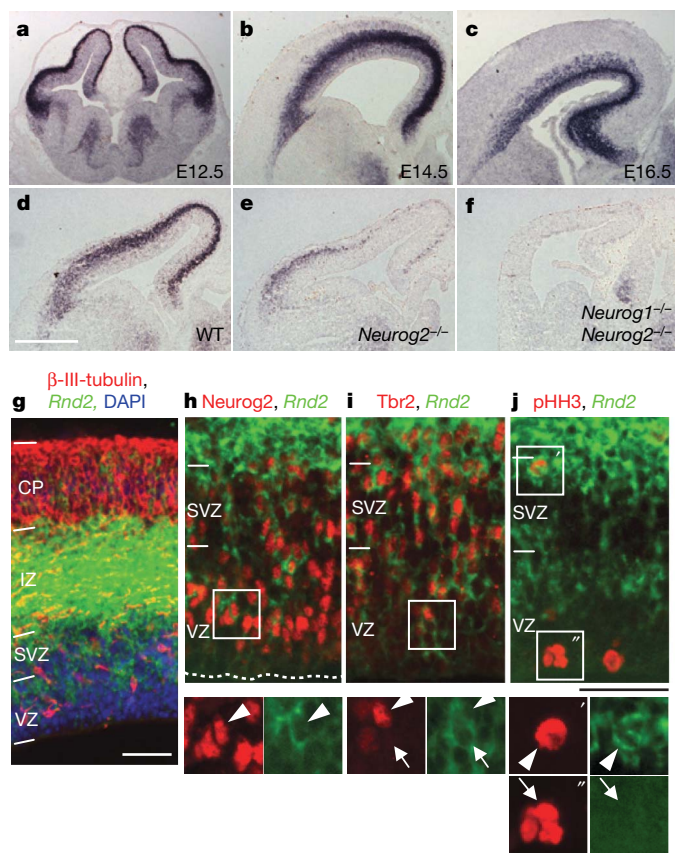


Figure 1 | *Rnd2* expression in the embryonic cerebral cortex requires *Neurog2*. **a–c**, *Rnd2* is expressed at high levels in the preplate of the cortex at E12.5 (**a**), and in the subventricular zone (SVZ) and intermediate zone (IZ) at E14.5 (**b**) and E16.5 (**c**). **d–f**, Cortical expression of *Rnd2* at E13.5 (**d**) was severely reduced in a *Neurog2* null-mutant embryo (**e**) and was abolished in a *Neurog1*;*Neurog2* double-mutant embryo (**f**). WT, wild type. **g**, Triple labelling for the neuronal marker β -III-tubulin (red), *Rnd2* transcripts (green) and the nuclear stain DAPI (blue) in the cortex at E14.5. CP, cortical plate; VZ, ventricular zone. **h, i**, Ventricular zone cells expressing *Rnd2* transcripts (green) also express *Neurog2* protein (red, **h**) and some of them express *Tbr2* protein (red, **i**). Double-labelled cells are indicated by arrowheads in higher-magnification panels, whereas a *Tbr2*⁺, *Rnd2*⁺ cell is shown by an arrow. **j**, *Rnd2* is expressed by intermediate progenitors that divide away from the ventricular surface (arrowhead in **j** points to a pHH3⁺, *Rnd2*⁺ cell) and not by ventricular zone progenitors dividing at the ventricular surface (arrow in **j** shows a pHH3⁺, *Rnd2*[−] cell). The dashed line in **h–j** marks the ventricular surface. Scale bars, 500 μ m (**a–f**), 50 μ m (**g–j**).

morphology (that is, harbouring at least three processes) from 48% to 70% (Fig. 2d–h). Moreover, the longest neurite of *Rnd2*-silenced neurons with a multipolar morphology was 69% longer than in control experiments (Fig. 2k, l), whereas neurons that remained unipolar or bipolar more frequently had a branched leading process (72% of *Rnd2*-silenced neurons versus 23% in controls) (Fig. 2i–n). Time-lapse imaging of electroporated slices additionally showed that the processes of *Rnd2*-silenced neurons were unstable (Fig. 2k, m and Supplementary Movies 1–4). A greater proportion of cells with multipolar morphologies were also observed among *Rnd2*-silenced cortical neurons maintained in dissociated cultures for 3 days (Supplementary Fig. 4k–n). Thus, *Rnd2* regulates both the shape and motility of migrating cortical neurons.

We next evaluated the contribution of *Rnd2* to the migration-promoting activity of the proneural gene *Neurog2*. Acute deletion of *Neurog2* in cortical neurons, obtained by electroporating the recombinase Cre in embryonic brains homozygous for a conditional mutant allele of *Neurog2* (*Neurog2*^{flax/flax}), resulted as expected in loss

of both *Neurog2* and *Rnd2* expression (Supplementary Fig. 10), and led to a block in radial migration similar to that observed in *Rnd2*-deficient neurons^{10,12} ($n = 3$; Fig. 3a, b). These observations suggested that *Neurog2* might promote the radial migration of cortical neurons through the induction of *Rnd2*. To address this possibility, we asked whether forced expression of *Rnd2* could rescue the radial migration defect of *Neurog2*-mutant neurons. Overexpression of *Rnd2* at high levels (that is, from the cytomegalovirus enhancer/ β -actin promoter) in the subventricular zone and intermediate zone perturbed the migration of wild-type cells and did not rescue the migration of cells in which *Neurog2* had been deleted (Supplementary Fig. 11, Supplementary Movies 5 and 6, and data not shown), suggesting that excessive levels of *Rnd2* are detrimental to cell migration (see also ref. 17). However, expression of *Rnd2* from the *NeuroD1* promoter (*pNeuroD1*), which is transiently and moderately active in newborn cortical neurons^{19,20} (F. Polleux, personal communication), resulted in a remarkable rescue of the radial migration defect of *Neurog2* mutant neurons ($n = 3$; Fig. 3a–c; Supplementary Fig. 13). Although only 28% of *Neurog2*-mutant cells (in *Neurog2*^{flax/flax} embryos electroporated with Cre) had migrated to the cortical plate after 4 days, 54% of *Neurog2*-mutant cells expressing *Rnd2* (in *Neurog2*^{flax/flax} embryos co-electroporated with Cre and the *pNeuroD1*-*Rnd2* plasmid) had reached the cortical plate, a proportion similar to that observed in a control experiment (52%), which indicates that *Rnd2* is a major effector of *Neurog2* for the promotion of radial migration. *Rnd2* expression did not restore the final phase of radial migration, as most *Neurog2*-mutant/*Rnd2*-expressing neurons failed to reach the upper layer of the cortical plate (Fig. 3c, d). This was not due to insufficient expression of *Rnd2* from the *NeuroD1* promoter in the cortical plate (Supplementary Fig. 4i), suggesting instead that another *Neurog2* target gene is required for correct neuronal positioning within the cortical plate.

The finding that *Rnd2* acts downstream of *Neurog2* to control the migration of neurons from the subventricular zone to the cortical plate, a late step in the neurogenic program, raised the question of whether *Neurog2* induces *Rnd2* expression directly or indirectly by a transcriptional cascade. Double-labelling experiments showed that *Rnd2* transcription is induced in ventricular zone cells that also express *Neurog2* but not necessarily the T-box protein *Tbr2*, the next transcription factor in the *Neurog2*-dependent transcriptional cascade activated during cortical neurogenesis^{2,21} (Fig. 1h–j). This suggests that *Neurog2*, rather than a factor further down the cascade, induces *Rnd2* expression in ventricular zone cells. To further address this possibility, we searched the vicinity of the *Rnd2* gene for evolutionary conserved regulatory elements that contain consensus binding sites for *Neurog2* (known as E-boxes^{8,22}). Using the University of California Santa Cruz Genome Browser (<http://www.genome.ucsc.edu/cgi-bin/hgGateway>), we identified a 366 base pair (bp) non-coding sequence located 3' to the *Rnd2* gene (named hereafter *Rnd2* 3' enhancer), which is highly conserved in mammalian genomes and contains two consensus *Neurog2*-type E-boxes (Fig. 4a and Supplementary Fig. 14). Using a transgenic mouse enhancer assay²³, we established that this element has transcriptional enhancer activity in the embryonic cortex ($n = 2$; Fig. 4b, c). Using a luciferase reporter assay in the embryonal carcinoma cell line P19 (ref. 23), we found that *Neurog2* efficiently activated transcription from the *Rnd2* 3' enhancer. This interaction required intact E-boxes, and it was specific for *Neurog2* because another proneural bHLH gene, *Ascl1*, had no activity in the same assay (Fig. 4d). Moreover, *Neurog2* was bound to the *Rnd2* 3' enhancer in cortical cells *in vivo*, as shown by chromatin immunoprecipitation using an antibody against *Neurog2* together with chromatin prepared from embryonic telencephalic tissue (Fig. 4e). Altogether, these experiments demonstrate that *Neurog2* induces *Rnd2* expression in the embryonic cortex by directly interacting with an enhancer located 3' to the gene (summary schemes in Supplementary Fig. 15a, b). Interestingly, induction of *Rnd2* expression by *Neurog2* did not require phosphorylation of *Neurog2* at

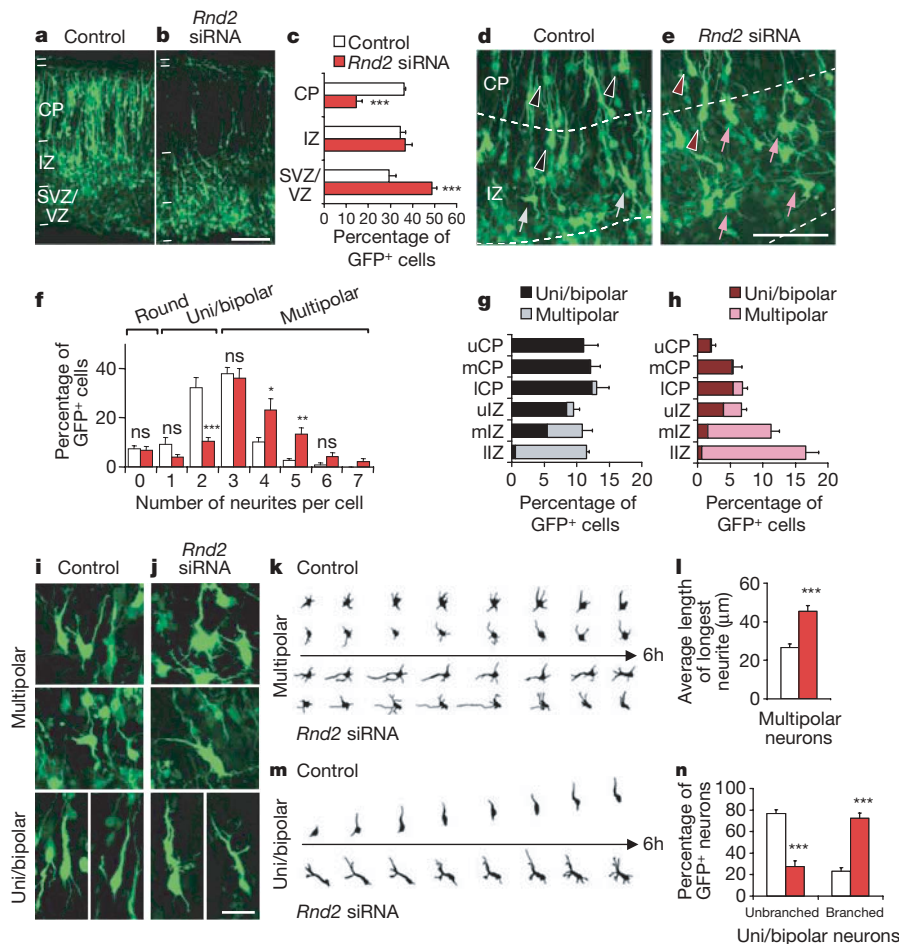


Figure 2 | *Rnd2* silencing blocks the radial migration of cortical projection neurons and alters their morphology. **a, b**, Migrating neurons in an E14.5 mouse brain electroporated with a control siRNA (**a**) or an siRNA against *Rnd2* (**b**) along with a green fluorescent protein (GFP)-expressing plasmid, sliced and cultivated for 4 days ($n = 6$, more than 3,000 cells per condition). Although control cells migrated into the cortical plate during the culture period, *Rnd2* siRNA-treated cells failed to reach the cortical plate and accumulated in the ventricular zone/subventricular zone. **d–n**, *Rnd2* knockdown affects the morphology of cortical neurons. In a control experiment (**d, f, g, i, k**), a large fraction of GFP⁺ neurons in the lower and median intermediate zone (lIZ and mIZ) were multipolar (grey arrows in **d**), whereas most neurons in the upper intermediate zone (uIZ) and in the lower, median and upper cortical plate (ICP, mCP and uCP) had acquired a unipolar or bipolar morphology (black arrowheads in **d**). After *Rnd2* knockdown (**e, f, h, j, k**), a larger fraction of neurons retained a multipolar morphology (pink arrows in **e**; uni/bipolar neurons identified with purple arrowheads) in the mIZ and uIZ. The graphs in **g** and **h** represent data from **d** and **e**, respectively, with colour-matched bars representing proportions of

uni/bipolar and multipolar neurons in control (black and grey) and *Rnd2* siRNA-treated (purple and pink) slices. **f**, *Rnd2* siRNA treatment increases the fraction of intermediate zone neurons with high numbers of neurites ($n \geq 300$ cells per condition). **i, j**, High-magnification pictures of neurons from **d** and **e**, respectively. **k, m**, Camera-lucida drawings of multipolar neurons (**k**) and uni/bipolar neurons (**m**) from control (top) and *Rnd2* siRNA-treated slices (bottom) analysed by time-lapse imaging for 6h. *Rnd2* deficient neurons had processes that were longer and more branched than wild-type neurons. The growth and retraction of the longest process of the multipolar *Rnd2*-deficient neurons were also accelerated (**k**). The uni/bipolar *Rnd2*-deficient neuron failed to migrate, in contrast to the wild-type uni/bipolar neuron, and its leading process was branched (**m**). **l**, Average length of the longest neurite in control and *Rnd2*-deficient multipolar neurons ($n = 25$; Mann–Whitney *U*-test). **n**, Fraction of control and *Rnd2*-deficient uni/bipolar neurons with a branched apical process ($n = 6$, more than 150 cells per condition). All graphs plot mean \pm s.e.m. Scale bars, 100 μ m (**a, b**), 20 μ m (**i, j**).

residue tyrosine241, a post-translational modification recently implicated in the regulation of cortical neuron migration by this proneural protein¹⁰ (Supplementary Fig. 16).

Neurog2 expression is rapidly downregulated after cortical progenitors have become post-mitotic¹⁰ (Fig. 1h), while the activity of the 3' enhancer and the expression of *Rnd2* are maintained in migrating neurons throughout the intermediate zone (Figs 1g and 4b). This suggests that factors other than Neurog2 are involved in maintaining the activity of the enhancer in migrating neurons. Several transcription factors of the T-box and bHLH families are thought to act in a transcriptional cascade downstream of Neurog2 (refs 2, 21). Among these, NeuroD2 (ref. 24) efficiently promoted transcription from the *Rnd2* 3' enhancer in the P19 luciferase assay, suggesting that this factor maintains *Rnd2* expression in the subventricular zone and intermediate zone by interacting with the same enhancer as

Neurog2. NeuroD1²⁵ and Tbr2²¹ also activated the 3' enhancer, albeit to a much lesser extent (Fig. 4d). Thus, the induction and maintenance of *Rnd2* expression in cortical neurons likely involves the sequential regulation of a single enhancer element by Neurog2 and several downstream transcription factors expressed at different stages of cortical neuron development (Supplementary Fig. 15c).

This study provides the first example to our knowledge of a developmental regulator controlling the spatio-temporal expression of a small GTP-binding protein. This is an important level of regulation because, unlike other Rho family members, Rnd proteins are constitutively active and are thought to be regulated primarily at the level of their expression²⁶. We also demonstrate that *Rnd2*, a regulator of cell morphology and cell movement, is an important component of the neurogenic program activated by *Neurog2* in the cerebral cortex. In the embryonic telencephalon, *Rnd2* is expressed at higher levels in

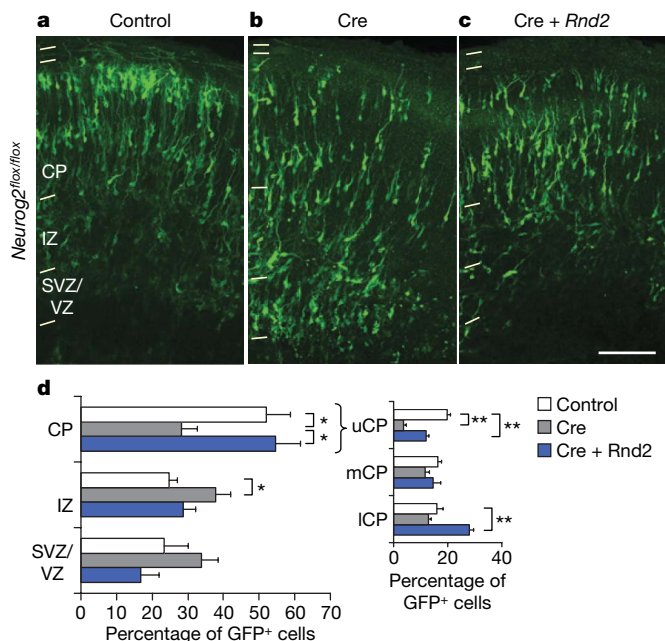


Figure 3 | Expression of *Rnd2* rescues the radial migration defect of *Neurog2*-mutant cortical neurons. **a, b**, Co-electroporation of the recombinase Cre with pNeuroD1–GFP (Supplementary Fig. 13) in E14.5 mouse brains homozygous for a conditional null mutant allele of *Neurog2* (*Neurog2*^{lox/lox}) resulted in an acute deletion of *Neurog2* and prevented many electroporated neurons from leaving the ventricular zone/subventricular zone and reaching the cortical plate (**b**), compared with control pNeuroD1–GFP electroporation alone (**a**). **c**, Co-electroporation of *Rnd2* expressed from the *NeuroD1* promoter with Cre and GFP largely rescued the migration defect caused by *Neurog2* deletion. Note that GFP⁺ cells remained scattered in the cortical plate rather than aligning in the upper part of the cortical plate as in (**a**), indicating that *Rnd2* expression did not rescue the last stage of cortical neuron migration within the cortical plate. **d**, Quantification of the distribution of GFP⁺ neurons in experiments shown in **a–c**. Cre + *Rnd2* electroporated cells left the ventricular zone/subventricular zone and reached the cortical plate as efficiently as control cells ($n = 3$, more than 1,100 cells per condition). However, further subdivision of the cortical plate showed that Cre + *Rnd2* electroporated cells accumulated mostly in the lower cortical plate in contrast with control cells that mostly reached the upper cortical plate. All graphs plot mean \pm s.e.m. Scale bar, 100 μ m.

dorsal than in ventral neurons, whereas the related gene *Rnd3* is expressed in a complementary manner with higher expression ventrally (our unpublished observations). Thus, *Rnd2* may be part of a program of neurogenesis that is specific to the dorsal telencephalon and confers unique properties to cortical neurons, whereas different *Rnd* proteins may participate in distinct programs of neurogenesis activated in other brain regions. *Rnd* proteins signal by multiple downstream pathways^{3,27,28} and may thus confer distinct migratory properties to different classes of newborn neurons.

METHODS SUMMARY

Analysis of *Rnd2* function. *Ex vivo* electroporation and organotypic slice culture were performed as described previously¹². Endotoxin-free plasmids were injected at $1 \mu\text{g} \mu\text{l}^{-1}$, whereas siRNA duplexes (Ambion) were injected at $10 \mu\text{M}$. Brain slices (300 μm) were cultured for up to four days *in vitro*, fixed with paraformaldehyde and processed for immunohistochemistry before image acquisition. Time-lapse movies were performed on electroporated brain slices that were cultured for 2 days *in vitro*, with images captured at 10-min intervals using an UltraVIEW Spinning Disk confocal microscope (Perkin Elmer) equipped with a Hamamatsu motorized stage. All graphs plot mean \pm s.e.m. statistics for dual comparisons and were generated using Student's *t*-tests unless specified, whereas statistics for multiple comparisons were generated using one-way analysis of variance followed by a suitable *post hoc* *t*-test (Supplementary Table 1); * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ for all statistics herein.

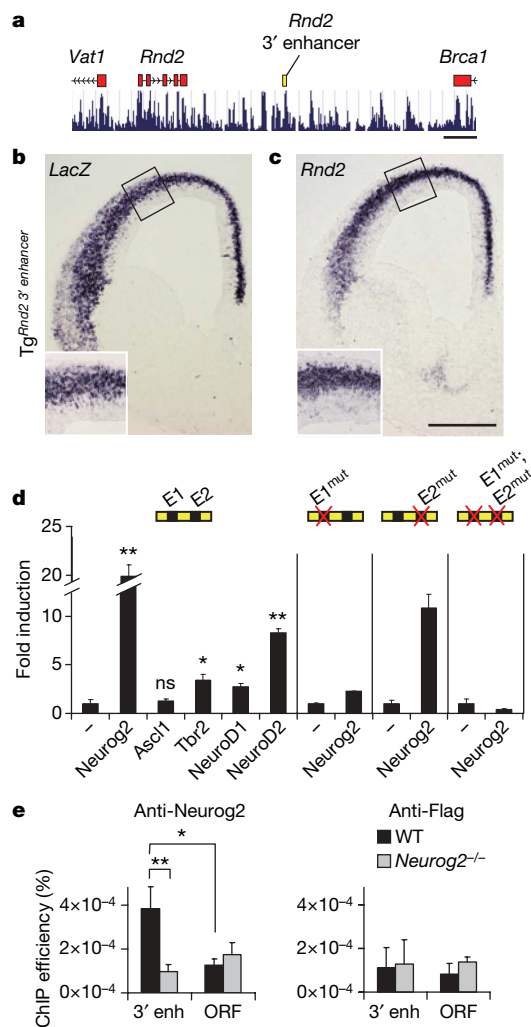


Figure 4 | *Neurog2* activates *Rnd2* expression in the cerebral cortex by direct binding to a 3' enhancer element. **a**, Identification of a conserved 366-bp element, named *Rnd2* 3' enhancer (yellow). **b, c**, This enhancer, linked to the minimal human β -globin promoter, drove expression of the *LacZ* reporter in the cortex of an E14.5 transgenic embryo in a pattern similar to that of endogenous *Rnd2* transcripts (**c**). **d**, *Neurog2* and other transcription factors of the embryonic cortex, including *Tbr2*, *NeuroD1* and *NeuroD2*, activated transcription from the *Rnd2* 3' enhancer in a luciferase reporter assay in P19 cells ($n = 3$), whereas the proneural protein *Ascl1* had no significant activity. Mutations of two consensus *Neurog2* binding motifs (*E1*^{mut} and *E2*^{mut}) reduced or abolished activation of the enhancer by *Neurog2*. **e**, Chromatin immunoprecipitation (ChIP) using an antibody against *Neurog2* and chromatin prepared from E13.5 telencephalon detected *Neurog2* binding to the *Rnd2* 3' enhancer *in vivo* ($n = 3$). Controls include ChIP with an anti-Flag antibody and with chromatin prepared from *Neurog2* mutant embryos, and detection of *Rnd2* open reading frame (ORF) in the immunoprecipitated material. All graphs plot mean \pm s.e.m. Scale bars, 2,000 bp (**a**), 500 μm (**b, c**).

Analysis of *Rnd2* regulation. *In situ* hybridization¹² was performed with an antisense RNA probe for *Rnd2* (prepared from IMAGE:5255562, GenBank accession number BI905297). The resultant nitroblue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP) signal was photographed then false coloured using Adobe Photoshop software. Immunohistochemistry was performed with standard protocols using the following primary antibodies: mouse polyclonal anti- β -tubulin (1:1000, Covance), mouse monoclonal anti-*Neurog2* (1:20, a gift from D. Anderson, California Institute of Technology), rabbit polyclonal anti-*Tbr2* (ref. 21), rabbit polyclonal anti-phospho histone H3 (1:500, Millipore), rabbit polyclonal anti-GFP (1:1000, Molecular Probes), mouse monoclonal anti-MAP2 (1:500, Sigma). Citrate antigen retrieval was performed for *Tbr2* immunostaining. Generation of transgenic animals, luciferase assays and chromatin immunoprecipitation experiments were performed as previously described²³.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 April; accepted 23 June 2008.

Published online 6 August 2008.

1. Fode, C. *et al.* A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes Dev.* **14**, 67–80 (2000).
2. Schuurmans, C. *et al.* Sequential phases of neocortical fate specification involve Neurogenin-dependent and -independent pathways. *EMBO J.* **23**, 2892–2902 (2004).
3. Chardin, P. Function and regulation of Rnd proteins. *Nature Rev. Mol. Cell Biol.* **7**, 54–62 (2006).
4. Marin, O. & Rubenstein, J. L. Cell migration in the forebrain. *Annu. Rev. Neurosci.* **26**, 441–483 (2003).
5. Ayala, R., Shu, T. & Tsai, L. H. Trekking across the brain: the journey of neuronal migration. *Cell* **128**, 29–43 (2007).
6. McManus, M. F. & Golden, J. A. Neuronal migration in developmental disorders. *J. Child Neurol.* **20**, 280–286 (2005).
7. Loturco, J. J. & Bai, J. The multipolar stage and disruptions in neuronal migration. *Trends Neurosci.* **29**, 407–413 (2006).
8. Bertrand, N., Castro, D. S. & Guillemot, F. Proneural genes and the specification of neural cell types. *Nature Rev. Neurosci.* **3**, 517–530 (2002).
9. Ross, S. E., Greenberg, M. E. & Stiles, C. D. Basic helix-loop-helix factors in cortical development. *Neuron* **39**, 13–25 (2003).
10. Hand, R. *et al.* Phosphorylation of Neurogenin2 specifies the migration properties and the dendritic morphology of pyramidal neurons in the neocortex. *Neuron* **48**, 45–62 (2005).
11. Ge, W. *et al.* Coupling of cell migration with neurogenesis by proneural bHLH factors. *Proc. Natl Acad. Sci. USA* **103**, 1319–1324 (2006).
12. Nguyen, L. *et al.* p27kip1 independently promotes neuronal differentiation and migration in the cerebral cortex. *Genes Dev.* **20**, 1511–1524 (2006).
13. Mattar, P. *et al.* A screen for downstream effectors of Neurogenin2 in the embryonic neocortex. *Dev. Biol.* **273**, 373–389 (2004).
14. Gohlke, J. M. *et al.* Characterization of the proneural gene regulatory network during mouse telencephalon development. *BMC Biol.* **6**, 15 (2008).
15. Nobes, C. D. *et al.* A new member of the Rho family, Rnd1, promotes disassembly of actin filament structures and loss of cell adhesion. *J. Cell Biol.* **141**, 187–197 (1998).
16. Negishi, M. & Katoh, H. Rho family GTPases and dendrite plasticity. *Neuroscientist* **11**, 187–191 (2005).
17. Nakamura, K. *et al.* *In vivo* function of Rnd2 in the development of neocortical pyramidal neurons. *Neurosci. Res.* **54**, 149–153 (2006).
18. Tabata, H. & Nakajima, K. Multipolar migration: the third mode of radial neuronal migration in the developing cerebral cortex. *J. Neurosci.* **23**, 9996–10001 (2003).
19. Huang, H. P. *et al.* Regulation of the pancreatic islet-specific gene BETA2 (neuroD) by neurogenin 3. *Mol. Cell. Biol.* **20**, 3292–3307 (2000).
20. Yokota, Y., Ring, C., Cheung, R., Pevny, L. & Anton, E. S. Nap1-regulated neuronal cytoskeletal dynamics is essential for the final differentiation of neurons in cerebral cortex. *Neuron* **54**, 429–445 (2007).
21. Englund, C. *et al.* Pax6, Tbr2, and Tbr1 are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *J. Neurosci.* **25**, 247–251 (2005).
22. Seo, S., Lim, J. W., Yellajoshyula, D., Chang, L. W. & Kroll, K. L. Neurogenin and NeuroD direct transcriptional targets and their regulatory enhancers. *EMBO J.* (2007).
23. Castro, D. S. *et al.* Proneural bHLH and Brn proteins co-regulate a neurogenic programme through cooperative binding to a conserved DNA motif. *Dev. Cell* **11**, 831–844 (2006).
24. Ince-Dunn, G. *et al.* Regulation of thalamocortical patterning and synaptic maturation by NeuroD2. *Neuron* **49**, 683–695 (2006).
25. Schwab, M. H. *et al.* Neuronal basic helix-loop-helix proteins (NEX and BETA2/Neuro D) regulate terminal granule cell differentiation in the hippocampus. *J. Neurosci.* **20**, 3714–3724 (2000).
26. Chardin, P. GTPase regulation: getting a Rnd Rock and Rho inhibition. *Curr. Biol.* **13**, R702–R704 (2003).
27. Kakimoto, T., Katoh, H. & Negishi, M. Identification of splicing variants of Rapostlin, a novel RND2 effector that interacts with neural Wiskott-Aldrich syndrome protein and induces neurite branching. *J. Biol. Chem.* **279**, 14104–14110 (2004).
28. Tanaka, H., Katoh, H. & Negishi, M. Pragmin, a novel effector of Rnd2 GTPase, stimulates RhoA activity. *J. Biol. Chem.* **281**, 10355–10364 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to S. Wood for technical assistance in generating transgenic embryos. We thank members of the Guillemot laboratory for suggestions on the manuscript, and F. Polleux (University of North Carolina) for the *pNeuroD1* plasmid and D. Anderson (California Institute of Technology) for the anti-Neurog2 antibody. J.I.-T.H. is supported by an Australian CJ Martin Fellowship (grant number 310616) and a Medical Research Council career development fellowship (UK), L.N. and H.W. by postdoctoral fellowships awarded by the European Molecular Biology Organization, and D.S.C. and C.Z. by Medical Research Council career development fellowships (UK). This work was supported by grants from the European Commission 6th Framework Programme for Research and Technological Development to F.G. and by institutional funds from the Medical Research Council (UK).

Author Contributions J.I.-T.H. and F.G. conceived all experiments, analysed data and wrote the manuscript. Together with J.I.-T.H., L.N. participated in *ex vivo* electroporation studies, H.W. conducted *in utero* electroporation experiments, C.Z. performed whole-embryo culture experiments, and D.S.C., D.S.-K. and J.-M.M. performed chromatin immunoprecipitation experiments. O.A. conducted the microarray screen, and F.B. and R.H. contributed to luciferase assays and immunostaining. All authors discussed results from experiments and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to F.G. (fguille@nimr.mrc.ac.uk).

METHODS

Animals. Mice were housed, bred and treated according to the guidelines approved by the UK Home Office under the Animal (Scientific Procedures) Act 1986. Protocols detailing the generation and genotyping of the genetically modified mice used in this article have been described previously for *Neurog1^{ΔLacZ}* (ref. 10), *Neurog2^{fllox}* (ref. 10), *Neurog2^{K1GFP}* (ref. 10), *Nes^{Cre}* (ref. 29).

Organotypic slice culture, *in situ* hybridization and imaging. Plasmids and siRNA duplexes were introduced by intraventricular injection of whole embryo heads using a Femtojet microinjector (Eppendorf) followed by electroporation using platinum electrodes (Nepagene) connected to an electroporator (ECM830, BTX). Experiments were performed only with slices (300 μm) of mid-rostral embryonic brain. *In utero* electroporation experiments were performed as described previously¹². Immunohistochemistry was performed with whole brain slices, or with slices that had been penetrated with 20% sucrose/PBS solution, embedded in optimum cutting temperature mounting medium and re-sectioned at 10 μm using a cryostat (Leica). The following antibodies were used for these studies: chicken polyclonal anti-GFP (1:700, Chemicon), mouse monoclonal anti-nestin (clone Rat-401, 1:100, Chemicon), rabbit polyclonal anti-human/mouse activated Caspase-3 (1:1,000, R&D Systems), mouse monoclonal anti-MAP2 (1:500, Sigma), rabbit polyclonal anti-Cre recombinase (1:1,000, Covance), rat monoclonal anti-BrdU (AbD Serotec). All fluorescent secondary antibodies were purchased from Invitrogen and used at a dilution of 1:800. For quantification, different subcompartments of the developing cortex (ventricular zone/subventricular zone, intermediate zone and cortical plate) were defined as described previously¹². Further subdivisions of the intermediate zone (lower-, medial- and upper) and cortical plate were delineated by an equal partitioning of each zone into three subcompartments. Images were acquired with an epifluorescent microscope (Axioplan 2, Zeiss) equipped with a CCD (charge-coupled device) digital camera (ProgRes C14, Jenoptik) and Openlab software (Improvision), or a laser scanning confocal microscope (Radiance 2100, BioRad). Cell counts were performed on representative fields using MetaMorph software (Molecular Devices). *In situ* hybridization was performed using standard protocols for *RhoA*¹¹, *Dcx*, *p35* (ref. 11) and *LacZ*²².

Plasmid constructs and siRNAs. The full-length coding sequence for mouse *Rnd2* was cloned by PCR using IMAGE:5503867 (GeneBank accession number BM461128) as template, then inserted into the *EcoRI* site of the pCIG2 vector to generate pCIG2-*Rnd2*-IRES-GFP. An amino-terminal Flag epitope was also inserted to generate a pCIG2-[Flag-*Rnd2*]-IRES-GFP expression construct. The p*NeuroD1*-*Rnd2*-IRES-GFP construct was generated by cloning the full-length *Rnd2* complementary DNA (cDNA) into the *EcoRI* fragment of p*NeuroD1*-IRES-GFP⁹. To generate an expression construct harbouring silent point mutations in the sequence recognized by *Rnd2* siRNA#1, full-length *Rnd2* cDNA was first amplified by PCR and cloned into pCR-TOPO (Stratagene). Then site-directed mutagenesis was performed using a QuikChange II Site-Directed Mutagenesis Kit (Stratagene) on a sequence-verified TOPO-*Rnd2* plasmid using the following primers: sense, 5'-GGAAATGAGGGCGAGATGCA-CAAAGACCGAGCCAAGAGCTGTA-3'; antisense, 5'-TACAGCTCTTGGC-TCGGTCTTTGTGCATCTCGCCCTCATTTCC-3'.

The underlined nucleotide (nt) residues on the sense strand identify silent mutations on nt651(T→C), nt654(G→A) and nt657(T→C) of the *Rnd2* cDNA sequence. After full sequence verification of a correctly mutagenized clone, the mutated *Rnd2* cDNA (hereby defined as *Rnd2**) was directionally cloned to generate pCIG2-[Flag-*Rnd2**]-IRES-GFP and p*NeuroD1*-*Rnd2**-IRES-GFP using *EcoRI* sites on both of these expression vectors. The Cre expression plasmid pCIG2-*Cre*-IRES-GFP as well as the *Neurog2^{Y241F}* expression plasmid have been described previously^{10,12}, whereas pCIG2-*Cre* was generated by digesting pCIG2-*Cre*-IRES-GFP with *NotI* and *PstI*, followed by blunting DNA ends with the Klenow fragment of DNA polymerase (Promega) and subsequent self-ligation.

The cDNA for an *Rnd2* variant protein harbouring a threonine 21→asparagine mutation (*Rnd2T21N*) was generated by site-directed

mutagenesis using a QuikChange II Site-Directed Mutagenesis Kit (Stratagene) and the following primer pair: forward, GCGGAGTGCGGCAAGAAC-GCGTTGCTGCAG; reverse, CTGCAGCAACGCGTTCTTGCCGACTCCGC.

For luciferase reporter assays, the *Rnd2* 3' enhancer was cloned from mouse genomic DNA by PCR into the *NheI/SalI* sites of the luciferase reporter vector p-βglo-Luc²². Mutations were generated using a QuikChange II Site-Directed Mutagenesis Kit (Stratagene) and the following primers:

To generate a mutation to the E1-binding site (designated as E1^{mut}): (forward) 5'-GCCTCTGCTGTTGACTCCTAAATAACAGTGATCTGTCTGCATATTAA-TGAGAT-3'; (reverse) 5'-ATCTCATTAATATGCAGACAGATCACTGTTA-TTTAGGAGTCAACAGCAGAGGC-3'.

To generate a mutation to the E2-binding site (designated as E2^{mut}): (forward) 5'-CACCAAAGGGAGGGGCGAGTGAGGAGTAGGGAAGGTGTA-3'; (reverse) 5'-TAACACCTTCCTACTCCTCACTGCCCTCCCTTTGGTG-3'.

An E1^{mut};E2^{mut} double mutant was generated by sequential site-directed mutagenesis using the reagents above. The *Rnd2* 3' enhancer sequence was also cloned as a 3' enhancer element into a *LacZ* reporter vector harbouring a minimal β-globin promoter for the subsequent generation of transgenic reporter mice²². All of the above-mentioned constructs were fully sequenced before their use in experiments.

The following siRNA duplexes (Ambion) were used in this study: Ambion siRNA ID#65909 (named *Rnd2* siRNA#1 in this study); (sense strand) 5'-GGGCGAGAUGCAUAAGGAUtt-3'; (antisense strand) 5'-AUCCUUAU-GCAUCUCGCCct-3'; Ambion siRNA ID#165812 (named *Rnd2* siRNA#2 in this study); (sense strand) 5'-GGAUCGACCAAGAGCUGUtt-3'; (antisense strand) 5'-ACAGCUCUUGGCUCAUCctt-3'.

The extent of *Rnd2* knockdown elicited by these *Rnd2* siRNAs was evaluated against the Silencer Negative Control #1 siRNA (ID#4611, Ambion) which encodes a 19 bp scrambled sequence with no significant homology to any known gene sequences from mouse, rat or human. siRNA sequences were cloned into the short hairpin RNA vector described previously^{12,30}.

Cell culture and western blots. Western blotting was performed using standard protocols²², and with the mouse embryocarcinoma cell line P19. The following primary antibodies were used for western blotting: mouse monoclonal anti-Flag (1:1,000, Sigma), mouse anti-actin C2 (1:1,000, Santa Cruz Biotechnology); and appropriate secondary antibodies: goat anti-rabbit IgG (H+L) horseradish peroxidase conjugate or goat anti-mouse IgG (H+L) horseradish peroxidase conjugate (1:5,000, Bio-Rad). Signal was revealed using enhanced chemiluminescent detection reagents according to the manufacturer's instructions (Amersham Biosciences). Dissociated cortical cells were prepared from electroporated brain slices, with 50,000 cells seeded into wells of a 24-well plate containing poly-L-lysine-coated glass coverslips and cultured for up to 3 days.

Chromatin immunoprecipitation assays. ChIP assays were performed with wild-type or *Neurog2*-mutant E13.5 dorsal telencephalic tissue, and with a monoclonal mouse anti-*Neurog2* antibody or a monoclonal mouse anti-Flag antibody (Sigma) as a negative control, as previously described²². Immunoprecipitated DNA was quantified using the iCycler iQ Real-Time PCR Detection System (BioRad) and an SYBR-Green-based kit for quantitative PCR (iQ Supermix, BioRad). Quantities of immunoprecipitated DNA were calculated by comparison with a standard curve generated by serial dilutions of input DNA. The primers used for amplification of the *Rnd2* 3' enhancer were: forward, 5'-TGCCTCTGCTGTGACTCCTAA-3'; reverse, 5'-CGGGTT-CATCCTGACACTGA-3'. Primers used for ORF were: forward: 5'-GGAGCCCTCGATGCTCTAGA-3'; reverse: 5'-AGACCTTAGGGAACCTT-CACCTTATAT-3'.

29. Goebbels, S. *et al.* Genetic targeting of principal neurons in neocortex and hippocampus of *NES-Cre* mice. *Genesis* **44**, 611–621 (2006).

30. Bron, R. *et al.* Boundary cap cells constrain spinal motor neuron somal migration at motor exit points by a semaphorin–plexin mechanism. *Neural Dev.* **2**, 21 (2007).

Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery

Libor Macůrek^{1*}, Arne Lindqvist^{1*}, Dan Lim², Michael A. Lampson³, Rob Klompmaaker¹, Raimundo Freire⁴, Christophe Clouin², Stephen S. Taylor⁵, Michael B. Yaffe² & René H. Medema¹

Polo-like kinase-1 (PLK1) is an essential mitotic kinase regulating multiple aspects of the cell division process¹. Activation of PLK1 requires phosphorylation of a conserved threonine residue (Thr 210) in the T-loop of the PLK1 kinase domain, but the kinase responsible for this has not yet been affirmatively identified^{2–6}. Here we show that in human cells PLK1 activation occurs several hours before entry into mitosis, and requires aurora A (AURKA, also known as STK6)-dependent phosphorylation of Thr 210. We find that aurora A can directly phosphorylate PLK1 on Thr 210, and that activity of aurora A towards PLK1 is greatly enhanced by Bora (also known as C13orf34 and FLJ22624), a known cofactor for aurora A (ref. 7). We show that Bora/aurora-A-dependent phosphorylation is a prerequisite for PLK1 to promote mitotic entry after a checkpoint-dependent arrest. Importantly, expression of a PLK1-T210D phospho-mimicking mutant partially overcomes the requirement for aurora A in checkpoint recovery. Taken together, these data demonstrate that the initial activation of PLK1 is a primary function of aurora A.

Cells arrested in G2 after activation of the DNA damage checkpoint can be stimulated to enter mitosis by addition of caffeine to the culture medium⁸. Caffeine inhibits ATM and ATR checkpoint kinases and switches off the checkpoint, allowing cells to recover, hence we refer to this process as checkpoint recovery⁸. Checkpoint recovery requires the presence of PLK1 (ref. 8), whereas mitotic entry in an unperturbed cell cycle can occur in the absence of PLK1 (refs 8, 9). Thus, checkpoint recovery provides us with a unique opportunity to study when and how PLK1 is activated.

To confirm that checkpoint recovery requires PLK1 kinase activity, rather than another function of PLK1, we used BI2536, a potent and selective inhibitor of the kinase activity of PLK1 (refs 9, 10). Checkpoint recovery of U2OS osteosarcoma cells was strongly inhibited when BI2536 was added together with caffeine (Fig. 1a), and mitotic entry was efficiently inhibited when BI2536 was added up to 3 h before mitosis (Fig. 1b), indicating that PLK1 kinase activity is indeed essential during recovery.

Next, we constructed stable U2OS-derived inducible cell lines using tetracycline-regulatable vectors for PLK1 that are insensitive to RNA interference (RNAi)¹¹. This allowed us to express PLK1 or the respective mutants in cells in which the endogenous PLK1 was depleted by short hairpin RNA (shRNA, Fig. 1c). Whereas wild-type PLK1 can revert the recovery defect of PLK1-depleted cells, expression of a kinase-inactive PLK1 mutant (K82R) failed to restore normal PLK1 function (Fig. 1d). Also, the non-phosphorylatable T210A mutant failed to restore checkpoint recovery, whereas the phospho-mimicking T210D mutant efficiently restored PLK1 function (Fig. 1d). Taken together, these data demonstrate the importance

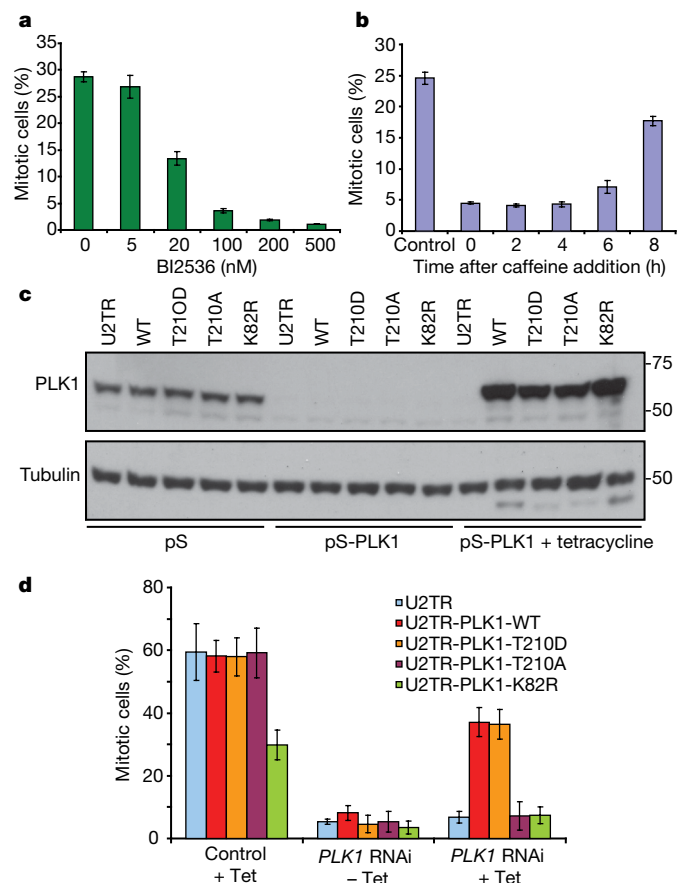


Figure 1 | Checkpoint recovery requires PLK1 kinase activity. Cells were arrested in G2 with doxorubicin as described in Methods, and subsequently induced to enter mitosis with caffeine. **a**, BI2536 was added to U2OS cells at the indicated concentrations together with caffeine. Cells were harvested after 9 h and mitotic entry was analysed by FACS. Error bars indicate standard deviation ($n = 3$). **b**, U2OS cells were treated with BI2536 (100 nM) at indicated time intervals after caffeine addition. Error bars indicate standard deviation ($n = 3$). **c**, **d**, U2OS-TR (U2TR), U2TR-PLK1-WT (wild type), -K82R (kinase dead), -T210D and -T210A cells were transfected with pS or pS-PLK1 together with pBabe-puro (**c**) or green fluorescent protein (GFP)-spectrin (**d**) and were selected with puromycin (**c**). Recovery from DNA damage and protein expression was induced by caffeine and tetracycline (Tet), respectively. Cell lysates were probed for PLK1 and tubulin expression on western blots (**c**). Alternatively, cells were collected after 9 h and mitotic entry was analysed by FACS (**d**). Error bars indicate standard deviation ($n = 6$).

¹Department of Medical Oncology and Cancer Genomics Centre, University Medical Center Utrecht, Utrecht 3584CG, The Netherlands. ²Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6069, USA. ⁴Unidad de Investigación, Hospital Universitario de Canarias, La Cuesta 38320, Spain. ⁵Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK.

*These authors contributed equally to this work.

of Thr 210 phosphorylation and kinase activation of PLK1 during recovery.

Next, we analysed Thr 210 phosphorylation by immunofluorescence using a phospho-specific antibody. We could detect Thr 210-phosphorylated PLK1 on centrosomes and kinetochores in early mitotic cells, in addition to a weaker but distinct staining of the chromatin (Fig. 2a). Depletion of PLK1 by RNAi completely removed the centrosome- and kinetochore-associated staining, whereas the chromatin-associated signal remained (Fig. 2a), indicating that a cross-reacting epitope is responsible for the latter. Centrosome-associated staining was not affected by treatment with BI2536 (Supplementary Fig. 1a), indicating that the antibody does not recognize another epitope generated by PLK1 activity. Using this antibody to monitor relative intensity of Thr 210-phosphorylated PLK1 on centrosomes, we found that Thr 210 phosphorylation can be observed in a subset of G2 cells with unseparated centrosomes, both in unperturbed cells and during checkpoint recovery (Fig. 2a and Supplementary Figs 1b, 2a). Thr 210 phosphorylation continues to accumulate over time, with a peak in (pro)metaphase, and gradually disappears from centrosomes during anaphase (Fig. 2a and Supplementary Figs 1b and 2a, b). Appearance of Thr 210 phosphorylation during caffeine-induced recovery coincided with phosphorylation of MYT1 on Thr 495, a previously characterized PLK1 site¹² (Fig. 2b). Thus, timing of PLK1 activation is closely correlated with Thr 210 phosphorylation under these conditions and occurs well before entry into mitosis.

To understand when exactly PLK1 kinase activity is first induced during an ongoing cell cycle, we made use of a fluorescence resonance energy transfer (FRET)-based indicator that allows real-time analysis of PLK1 activation in live cells (Supplementary Fig. 3). This indicator was modelled on a FRET-based biosensor previously used for other kinases^{13,14}, and was adapted for PLK1 by insertion of a consensus motif for PLK1 phosphorylation in the linker region between the donor and acceptor fluorophore¹⁵. The resulting FRET probe can be directly phosphorylated on this site by PLK1 *in vitro* (Supplementary Fig. 3c), and *in vivo* this site is phosphorylated in mitosis, at a time when PLK1 activity is high (Supplementary Fig. 3d, e). Phosphorylation of this PLK1 site is expected to reduce FRET efficiency between donor and acceptor fluorophores, allowing us to use this as an indirect measure for PLK1 kinase activity in the cell. Indeed, we detected a clear shift in FRET efficiency at 4 to 6 h before the onset of mitosis (Fig. 2c, d and Supplementary Fig. 3f). FRET efficiency was further reduced as cells got closer to mitosis, with a maximal shift during prometaphase and metaphase (Fig. 2c). This change in FRET efficiency was not observed when the phospho-site in the consensus motif was mutated to alanine, indicating that phosphorylation is essential to produce the shift in FRET efficiency (Supplementary Fig. 3g).

The change in FRET efficiency in late G2 cells was completely lost on depletion of PLK1 (Fig. 2c, d and Supplementary Fig. 3f, h) or after treatment with BI2536 (Supplementary Fig. 3i), indicating that the change requires PLK1 activation. Both PLK1-depleted as well as

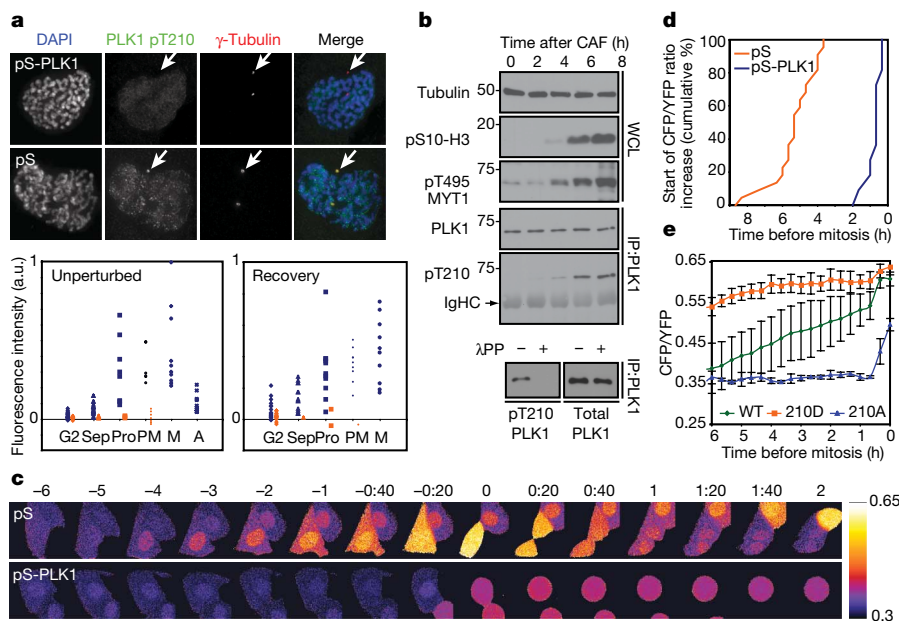


Figure 2 | PLK1 activation occurs several hours before mitosis and requires Thr 210 phosphorylation. **a**, Example (arrows) and quantification (plots) of centrosomal PLK1 Thr 210 phosphorylation during normal mitotic entry (left) and checkpoint recovery, fixed 4 h after caffeine addition (right). Each dot corresponds to a single cell. Blue dots, untreated cells; orange dots, pS-PLK1 transfected cells. G2, cells with two centrosomes; Sep, separated centrosomes; Pro, condensed DNA; PM, prometaphase; M, metaphase; A, anaphase; a.u., arbitrary units. Whereas the pS-PLK1 and control cells grew on the same coverslip, the unperturbed and recovery situation were on different coverslips. The fluorescence intensities are thus comparable within each graph but can only be approximatively compared between the graphs. For number of cells analysed and normalization versus γ -tubulin fluorescence, see Supplementary Fig. 2a. **b**, Timing of pT210 modification of PLK1 in recovery. U2OS cells treated as described in Fig. 1, collected at indicated time intervals after caffeine (CAF) addition and subjected to immunoprecipitation with anti-PLK1 antibody. Immunoprecipitates or whole-cell lysates (WCL) were probed with the indicated antibodies ($n = 3$). The arrow indicates position of the immunoglobulin used for

immunoprecipitation. Alternatively, the immunoprecipitate was treated with λ -phosphatase (λ PP) before separation on SDS-PAGE. **c**, Time-lapse sequence showing the false-coloured cyano fluorescent protein/yellow fluorescent protein (CFP/YFP) emission ratio of a U2OS cell expressing a FRET probe for PLK1 activity. Top panel, control transfection; lower panel, pS-PLK1 transfection. The time relative to mitosis is indicated above the images (hour:min). **d**, Timing of PLK1 activation in 23 pS-transfected cells or 12 pS-PLK1-transfected cells. The graph shows the cumulative percentage of the start of the CFP/YFP ratio increase, defined as the time point after the last negative derivative of the nuclear CFP/YFP ratio before mitosis ($n = 2$). **e**, Average and standard deviation of PLK1 FRET probe CFP/YFP emission ratio in pS-PLK1 transfected cells, reconstituted with tetracycline-induced expression of RNAi-insensitive PLK1. Whereas reconstitution with wild-type PLK1 (green lines $n = 5$) restores timing and appearance of the shift in CFP/YFP ratio, PLK1-T210A (blue lines $n = 3$) fails to induce a CFP/YFP change in G2, and PLK1-T210D (orange lines $n = 5$) induces a CFP/YFP ratio change at more than 6 h before mitosis ($n = 2$).

BI2536-treated cells displayed a partial shift in FRET signal at the later time points (Fig. 2c, d and Supplementary Fig. 3f, i), suggesting that the indicator can be phosphorylated by other kinases in mitosis. Nonetheless, our data clearly show that the FRET indicator specifically reads out PLK1 activity in G2. The shift in FRET efficiency also occurred roughly 5 h before mitosis in G2-arrested cultures induced to enter mitosis by the addition of caffeine (Supplementary Fig. 3j), indicating that PLK1 activation occurs at a similar stage in G2 during unperturbed growth and checkpoint recovery. Most importantly, reconstitution of PLK1-depleted cells with the T210D mutant already produced full PLK1 activation at a time in G2 when wild-type PLK1 was first activated, whereas expression of the T210A mutant failed to produce PLK1-specific activity (Fig. 2e). This shows that Thr 210 phosphorylation is required for activation of PLK1 *in vivo* and indicates that Thr 210 phosphorylation could be sufficient for initial activation.

Phosphorylation of Ser 137 has also been suggested to activate PLK1 (refs 2, 16). Interestingly, the amino acid sequences surrounding the Ser 137 and Thr 210 sites are highly similar suggesting that both sites could be phosphorylated by the same kinase. However, expression of a S137A/T210D double mutant also produced full PLK1 activation early in G2, similar to what we observed with the T210D mutant (Supplementary Fig. 4). This demonstrates that phosphorylation of Ser 137 is not required for initial PLK1 activation, and is consistent with earlier reports that Ser 137 phosphorylation is not observed before anaphase^{2,17}.

Thr 210 lies within a consensus site for aurora kinases. Interestingly, aurora A and PLK1 are both at centrosomes at a time when Thr 210 phosphorylation of PLK1 is observed (Supplementary Fig. 1c), their interaction is induced during recovery (Supplementary Fig. 5a), both aurora A and PLK1 are inhibited in response to DNA damage^{18,19}, and

overexpression of aurora A abrogates the G2 checkpoint after DNA damage^{18,20}. Therefore, we considered aurora A as the primary kinase for phosphorylation of Thr 210. Indeed, co-expression of a kinase-dead mutant of PLK1 and kinase-active aurora A in *Escherichia coli* led to phosphorylation of the Thr 210 site (Fig. 3a). In addition, recombinant aurora A could phosphorylate PLK1 on the Thr 210 site *in vitro*, albeit at a low efficiency (Fig. 3b). Activity of recombinant aurora A towards PLK1 was greatly enhanced by the addition of recombinant Bora (Fig. 3c), a previously described co-activator of aurora A⁷ that was shown to interact with PLK1 (refs 21, 22). Aurora A immunoprecipitated from asynchronous cultures failed to phosphorylate PLK1 directly, but inclusion of recombinant Bora resulted in very efficient phosphorylation of PLK1 at Thr 210 (Fig. 3d). Also aurora A immunoprecipitated from cells synchronized in G2, when aurora A and Bora form a complex^{21,22} (Supplementary Fig. 5b, c), readily phosphorylated PLK1 at Thr 210 (Supplementary Fig. 5d). Thus, aurora A can directly phosphorylate PLK1 on Thr 210, and this is facilitated by Bora. Indeed, we could detect Thr 210-phosphorylated PLK1 in complex with Bora as early as 4 h after caffeine addition in checkpoint recovery (Supplementary Fig. 5e).

Importantly, removal of aurora A by RNAi (Supplementary Fig. 6a, b) interfered with phosphorylation of Thr 210 in U2OS cells (Fig. 3e and Supplementary Fig. 7a), as well as in HeLa cells (data not shown), and abolished activation of PLK1 in G2 cells (Fig. 3f). Similarly, inhibition of aurora A with MLN8054 (a selective inhibitor for aurora A)²³ interfered with activation of PLK1 in G2, whereas treatment with ZM447439 (a selective inhibitor for aurora B), failed to do so (Supplementary Fig. 7b). This indicates that aurora A, but not aurora B, is responsible for the activation of PLK1 in G2. Indeed, removal of aurora A by two independent pS-aurora A targeting vectors caused a prominent inhibition in checkpoint recovery

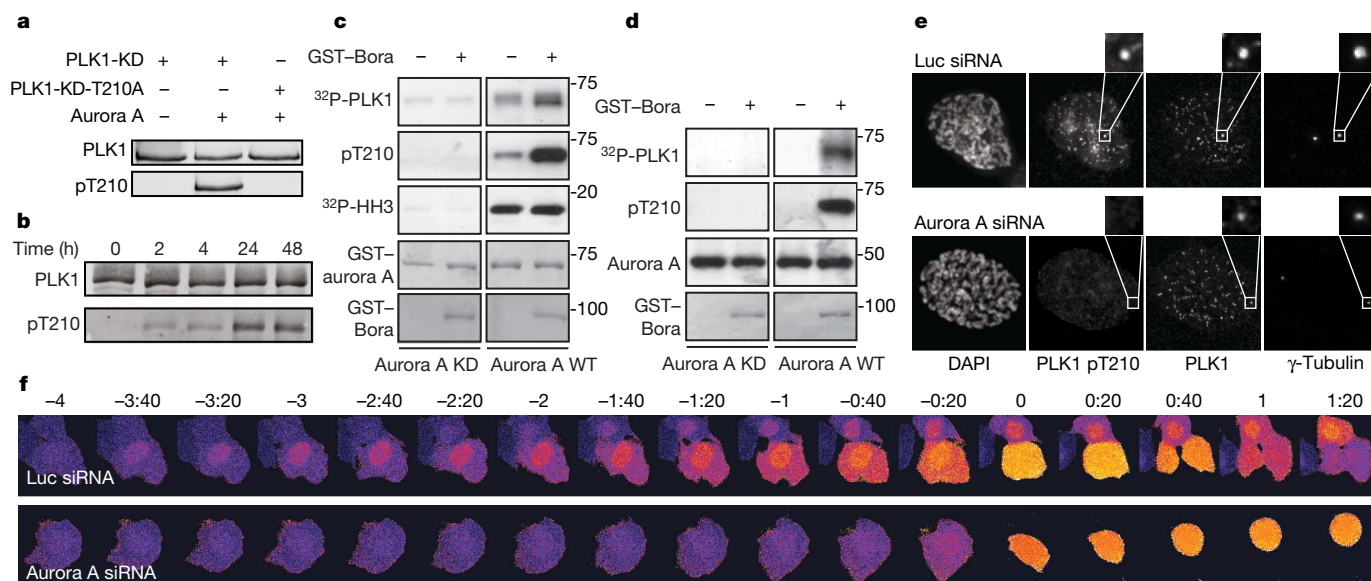


Figure 3 | PLK1 activation depends on aurora A. **a**, PLK1 is a direct substrate for aurora A kinase *in vitro*. Kinase-inactive (D176N) constructs of PLK1 with or without a T210A mutation were expressed in *E. coli* individually or co-expressed with aurora A, and Thr 210 phosphorylation was determined on a western blot. **b**, Aurora A can phosphorylate PLK1 Thr 210 *in vitro*. Glutathione S-transferase (GST)–aurora A and kinase-dead maltose-binding protein (MBP)–PLK1 were purified from *E. coli* and incubated in kinase buffer. Aliquots taken at the indicated time points (h) were probed for PLK1 Thr 210 phosphorylation. **c**, Aurora A phosphorylation of PLK1 is enhanced by Bora *in vitro*. GST–aurora A (wild-type or kinase-dead) and kinase-dead His–PLK1 were purified from *E. coli* and incubated for 30 min in kinase buffer with or without recombinant GST–Bora in the presence of ³²P-γ-ATP. Blots were probed with anti-pT210. Phosphorylation of PLK1 and histone H3 was analysed by autoradiography. The amount of incorporated ³²P in PLK1 and histone H3 was comparable in

these assays, and ~20% of PLK1-KD (kinase dead) was phosphorylated as determined from the amount of incorporated ³²P. **d**, Aurora A (wild-type and kinase dead) immunoprecipitated from HEK293 cells²⁷ was incubated in kinase buffer in the presence of ³²P-γ-ATP for 30 min with kinase-dead His-tagged PLK1 purified from *E. coli*. Recombinant GST–Bora was added where indicated. **e**, Aurora A RNAi prevents Thr 210 phosphorylation of centrosomal PLK1. U2OS cells were treated with luciferase (Luc) or aurora A RNAi for 48 h, and were fixed and processed for immunofluorescence with the indicated antibodies. Inserts show magnification of one of the centrosomes. Note that a part of the nuclear pT210 staining is unspecific. **f**, Aurora A RNAi affects G2 activation of PLK1. Images show time-lapse with 20 min intervals of false-coloured CFP/YFP emission ratios of a luciferase (top) and an aurora A (bottom) small interfering RNA (siRNA)-transfected cell entering mitosis (*n* = 3).

induced by caffeine (Fig. 4a and Supplementary Fig. 8a). Similarly, using MLN8054 we observed a clear reduction in recovery (Fig. 4b and Supplementary Fig. 8a), whereas selective inhibition of aurora B by ZM447439 did not interfere with checkpoint recovery (Supplementary Fig. 8b). Moreover, depletion of aurora A blocked the appearance of phosphorylation of Thr210 in normal G2 (Supplementary Fig. 9a) and during recovery (Fig. 4d), indicating that aurora A promotes recovery through phosphorylation of Thr210, leading to consequent activation of PLK1.

Because *in vitro* phosphorylation of PLK1 by aurora A was facilitated by addition of Bora, we also tested whether Bora is required for initial PLK1 activation in G2. Depletion of Bora by two independent shRNAs (Supplementary Fig. 6c–f) led to a prominent block in caffeine-induced checkpoint recovery (Fig. 4c), coincident with loss of Thr210 phosphorylation (Fig. 4d), indicating that Bora is required for aurora-A-dependent activation of PLK1 in G2. The requirement for Bora and aurora A was not restricted to caffeine-induced recovery, but was also observed in cells that were allowed to recover spontaneously (Supplementary Fig. 8c). No effect on recovery and Thr210 phosphorylation was observed after depletion of TPX2 (Fig. 4c, d and Supplementary Fig. 6g), another co-activator of aurora A²⁴.

To confirm that the function of aurora A in checkpoint recovery is indeed executed in part by means of phosphorylation of the Thr210

site, we investigated if the recovery defect induced by depletion of aurora A could be overcome by expression of the T210D mutant. Interestingly, expression of the T210D mutant of PLK1 partially abolished the requirement of aurora A for checkpoint recovery (Fig. 4e), consistent with our conclusion that PLK1 activation in G2 is controlled by means of Thr210 phosphorylation by aurora A. The fact that the recovery defect was rescued only partially by expression of the T210D mutant suggests that aurora A can phosphorylate additional components of the recovery pathway. In fact, it is known that aurora A can phosphorylate CDC25B (ref. 25), another essential regulator of checkpoint recovery⁸.

Taken together, our data show that activation of PLK1 first occurs on average 5 h before the onset of mitosis. This event crucially depends on aurora-A-dependent phosphorylation of Thr210 on PLK1, both during an unperturbed cell cycle and during checkpoint recovery. This function of aurora A becomes essential during checkpoint recovery, consistent with the unique role for PLK1 in this process. Because both PLK1 and aurora A are recruited to centrosomes in G2, it is tempting to speculate that this activation will first occur on centrosomes. Using our FRET probe, changes in FRET are first observed in the nucleus, but this cannot be taken as evidence that PLK1 activation first occurs in the nucleus. For one, the phosphatase that can dephosphorylate the FRET probe could be restricted to the cytoplasm in G2. Alternatively, small amounts of PLK1 activated at the centrosome might not be detected with our experimental design because the FRET probe can freely diffuse within the cytoplasm. Also, PLK1 that is activated at centrosomes could immediately translocate to the nucleus. Nonetheless, although our current data do not allow us to conclude where PLK1 is activated, they clearly demonstrate that aurora A is the upstream kinase for PLK1, which, together with Bora, controls activation of PLK1 in G2. Interestingly, PLK1 was recently shown to regulate aurora A through Bora^{21,22}, indicating that a feed-forward mechanism may promote efficient activation of both kinases in G2.

Our data show that aurora A, but not aurora B, is required for the initial activation of PLK1 in G2. However, we cannot exclude that aurora B phosphorylates PLK1 on the Thr210 site at later stages during mitosis. Interestingly, cross-talk between PLK1 and aurora B was shown to occur at centromeres, a predominant site of PLK1 and aurora B recruitment during (pro)metaphase, where both act to regulate chromosome alignment²⁶. Later, both proteins co-localize on the spindle midzone and control cytokinesis. Thus, it will be interesting to test if aurora B promotes Thr210 phosphorylation at later stages of mitosis, specifically at sites where PLK1 function is most crucial.

METHODS SUMMARY

Checkpoint recovery was analysed in G2-arrested cultures induced to re-enter the cell cycle by addition of caffeine, or on spontaneous recovery. Requirement of aurora A and Bora in recovery was analysed by specific depletion by shRNA, and recovery was monitored by FACS-based determination of MPM2- or H3-positive (mitotic) cells. Phosphorylation of PLK1 at Thr210 was determined by immunofluorescence and western blotting. In parallel, PLK1 activation in live cells was determined using a FRET-based biosensor.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 May; accepted 23 June 2008.

Published online 9 July 2008; corrected 5 August (details online).

- van Vugt, M. A., Bras, A. & Medema, R. H. Restarting the cell cycle when the checkpoint comes to a halt. *Cancer Res.* **65**, 7037–7040 (2005).
- Jang, Y. J., Ma, S., Terada, Y. & Erikson, R. L. Phosphorylation of threonine 210 and the role of serine 137 in the regulation of mammalian polo-like kinase. *J. Biol. Chem.* **277**, 44115–44120 (2002).
- Kelm, O., Wind, M., Lehmann, W. D. & Nigg, E. A. Cell cycle-regulated phosphorylation of the *Xenopus* polo-like kinase Plx1. *J. Biol. Chem.* **277**, 25247–25256 (2002).
- Ohkura, H., Hagan, I. M. & Glover, D. M. The conserved *Schizosaccharomyces pombe* kinase plo1, required to form a bipolar spindle, the actin ring, and septum, can drive septum formation in G1 and G2 cells. *Genes Dev.* **9**, 1059–1073 (1995).

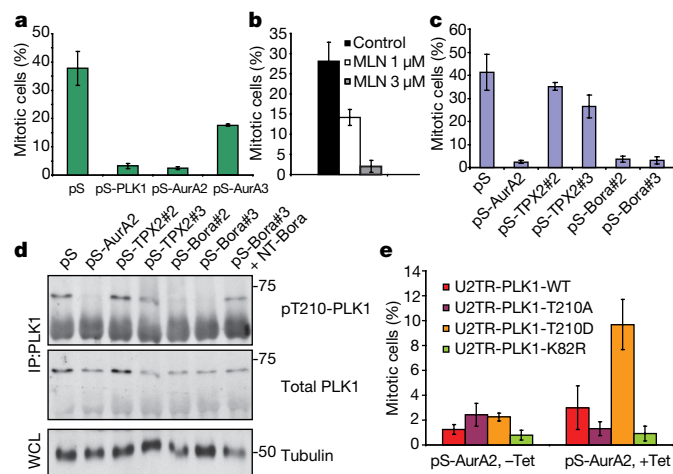


Figure 4 | Aurora-A-dependent phosphorylation of PLK1 is required for recovery. **a**, U2OS cells were co-transfected with the indicated pSuper constructs together with GFP-spectrin, and were treated as described in Fig. 1. Cells were harvested at 9 h after addition of caffeine, and mitotic entry was analysed by FACS using anti-MPM2. Error bars indicate standard deviation ($n = 4$). **b**, U2OS cells were treated as described in Fig. 1, and caffeine-induced recovery was performed in the presence or absence of 1 μ M or 3 μ M MLN8054. Cells were harvested at 9 h after addition of caffeine, and mitotic entry was analysed by FACS using anti-MPM2. Error bars indicate standard deviation ($n = 3$). **c**, U2OS cells were transfected with constructs targeting Bora or TPX2 (Supplementary Fig. 6) in combination with GFP-spectrin. Forty-eight hours after transfection, cells were treated as described in Fig. 1. Cells were harvested at 9 h after addition of caffeine, and mitotic entry was analysed by FACS using anti-MPM2. Error bars indicate standard deviation ($n = 3$). **d**, Depletion of aurora A or Bora prevents Thr210 phosphorylation during recovery. U2OS cells were transfected with the indicated pSuper-constructs in combination with pBabe-puro. Cells were selected with puromycin and treated as described in Fig. 1. Where indicated, non-targetable Bora was co-transfected. Cell lysates were prepared at 9 h after addition of caffeine, and PLK1 was immunoprecipitated. Thr210 phosphorylation of PLK1 was determined by immunoblotting. **e**, Induction of PLK1-T210D expression but not PLK1-WT partially rescues loss of aurora A in recovery. U2TR-PLK1-WT, -K82R, -T210D and -T210A cell lines were transfected with ps or ps-aurora A together with GFP-spectrin. Cells were treated as described in Fig. 1. Recovery from DNA damage and protein expression was induced by caffeine and tetracycline, respectively. Cells were collected after 9 h, and mitotic entry was analysed by FACS. Error bars indicate standard deviation ($n = 5$).

5. Okano-Uchida, T. *et al.* Distinct regulators for Plk1 activation in starfish meiotic and early embryonic cycles. *EMBO J.* **22**, 5633–5642 (2003).
6. Qian, Y. W., Erikson, E. & Maller, J. L. Purification and cloning of a protein kinase that phosphorylates and activates the polo-like kinase Plx1. *Science* **282**, 1701–1704 (1998).
7. Hutterer, A. *et al.* Mitotic activation of the kinase Aurora-A requires its binding partner Bora. *Dev. Cell* **11**, 147–157 (2006).
8. van Vugt, M. A., Bras, A. & Medema, R. H. Polo-like kinase-1 controls recovery from a G2 DNA damage-induced arrest in mammalian cells. *Mol. Cell* **15**, 799–811 (2004).
9. Lenart, P. *et al.* The small-molecule inhibitor BI 2536 reveals novel insights into mitotic roles of polo-like kinase 1. *Curr. Biol.* **17**, 304–315 (2007).
10. Steegmaier, M. *et al.* BI 2536, a potent and selective inhibitor of polo-like kinase 1, inhibits tumor growth *in vivo*. *Curr. Biol.* **17**, 316–322 (2007).
11. van Vugt, M. A. *et al.* Polo-like kinase-1 is required for bipolar spindle formation but is dispensable for anaphase promoting complex/Cdc20 activation and initiation of cytokinesis. *J. Biol. Chem.* **279**, 36841–36854 (2004).
12. Nakajima, H., Toyoshima-Morimoto, F., Taniguchi, E. & Nishida, E. Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. *J. Biol. Chem.* **278**, 25277–25280 (2003).
13. Violin, J. D., Zhang, J., Tsien, R. Y. & Newton, A. C. A genetically encoded fluorescent reporter reveals oscillatory phosphorylation by protein kinase C. *J. Cell Biol.* **161**, 899–909 (2003).
14. Zhang, J., Hupfeld, C. J., Taylor, S. S., Olefsky, J. M. & Tsien, R. Y. Insulin disrupts β -adrenergic signalling to protein kinase A in adipocytes. *Nature* **437**, 569–573 (2005).
15. Fuller, B. G. *et al.* Midzone activation of aurora B in anaphase produces an intracellular phosphorylation gradient. *Nature* **453**, 1132–1136 (2008).
16. Qian, Y. W., Erikson, E. & Maller, J. L. Mitotic effects of a constitutively active mutant of the *Xenopus* polo-like kinase Plx1. *Mol. Cell. Biol.* **19**, 8625–8632 (1999).
17. van de Weerd, B. C. *et al.* Uncoupling anaphase-promoting complex/cyclosome activity from spindle assembly checkpoint control by deregulating polo-like kinase 1. *Mol. Cell. Biol.* **25**, 2031–2044 (2005).
18. Krystyniak, A., Garcia-Echeverria, C., Prigent, C. & Ferrari, S. Inhibition of Aurora A in response to DNA damage. *Oncogene* **25**, 338–348 (2006).
19. Smits, V. A. *et al.* Polo-like kinase-1 is a target of the DNA damage checkpoint. *Nature Cell Biol.* **2**, 672–676 (2000).
20. Marumoto, T. *et al.* Roles of aurora-A kinase in mitotic entry and G2 checkpoint in mammalian cells. *Genes Cells* **7**, 1173–1182 (2002).
21. Chan, E., Santamaria, A., Sillje, H. & Nigg, E. Plk1 regulates mitotic Aurora A function through β TrCP-dependent degradation of hBora. *Chomosome* (in the press).
22. Seki, A. *et al.* Plk1- and β -TrCP-dependent degradation of Bora controls mitotic progression. *J. Cell Biol.* **181**, 65–78 (2008).
23. Manfredi, M. G. *et al.* Antitumor activity of MLN8054, an orally active small-molecule inhibitor of Aurora A kinase. *Proc. Natl Acad. Sci. USA* **104**, 4106–4111 (2007).
24. Eyers, P. A., Erikson, E., Chen, L. G. & Maller, J. L. A novel mechanism for activation of the protein kinase Aurora A. *Curr. Biol.* **13**, 691–697 (2003).
25. Dutertre, S. *et al.* Phosphorylation of CDC25B by Aurora-A at the centrosome contributes to the G2–M transition. *J. Cell Sci.* **117**, 2523–2531 (2004).
26. Rosasco-Nitcher, S. E., Lan, W., Khorasanizadeh, S. & Stukenberg, P. T. Centromeric Aurora-B activation requires TD-60, microtubules, and substrate priming phosphorylation. *Science* **319**, 469–472 (2008).
27. Girdler, F. *et al.* Validating Aurora B as an anti-cancer drug target. *J. Cell Sci.* **119**, 3664–3675 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are thankful to B. Ingelheim for the gift of BI2536, Millennium Pharmaceuticals for MLN8054, E. Nigg for antibodies and sharing unpublished work, A. Desai and I. Cheeseman for providing the pLC113 plasmid, M. Tanenbaum for U2OS cells expressing YFP-tubulin, S. Sen for GST-aurora A constructs, M. Vromans for recombinant GST-aurora A, Q. Verstralen for technical assistance, D. Littler for active recombinant PLK1, P. de Graaf for the U2OS-TR cell line, G. Kops for making the Deltavision imaging system available, L. Kleij for technical assistance with confocal microscopy, and all members of the Medema laboratory for discussions. We thank M. Alvarez, S. Lens, G. Kops and M. Tanenbaum for critically reading the manuscript. This work was supported by the Dutch Cancer Society (L.M. UU2006-3579), the Dutch Organisation for Scientific Research (R.H.M. ZonMw 918.46.616 and A.L. 916.86.083), the Association for International Cancer Research (L.M. 08-0172) and the NIH (M.B.Y. GM-60594 and CA112967). A.L. was also funded by the Wenner-Gren foundations. R.H.M. was funded by the Netherlands Genomic Initiative of the Netherlands Organisation for Scientific Research (NWO).

Author Contributions L.M., A.L., D.L., M.A.L., C.C., R.K., R.F. and R.H.M. performed experimental work. All authors were involved in project planning and manuscript preparation.

Author Information The coding sequence for full-length Bora was deposited in the GenBank database (EU834129). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.H.M. (r.h.medema@umcutrecht.nl).

METHODS

Plasmids, siRNAs, antibodies and reagents. pSuper-PLK1 (ref. 8), GFP-spectrin (ref. 8), pBabePuro (ref. 28), GST-AurA-WT and GST-AurA-KD (ref. 29), and plasmids encoding RNAi-non-targetable Myc-tagged human PLK1-WT as well as its K82R, T210D and T210A mutants¹⁷ have been described previously. Fragments corresponding to PLK1 were subcloned into HindIII/XbaI sites of pcDNA4/TO (Invitrogen) and PLK1-WT was cloned into pC113 plasmid in frame with the EGFP-TEV-S tag (referred to as PLK-LAP)³⁰. siRNAs targeting luciferase (as a negative control) and aurora A (GGCAACCACTGTACCTCAT) were from Ambion. shRNA plasmids targeting two distinct sites in aurora A (GGCAACCACTGTACCTCAT, ATGCCCTGCTTACTGTCA³¹), Bora (GGT TGATCCTATAGAGATA, GTGAAGATGAGGAAGATAA) and for TPX2 (GGG CAAAACTCCTTTGAGA, GGATGAACACTTTGAATTT) were generated by cloning of corresponding oligonucleotides into pSuper plasmid²⁸. Full-length Bora (ENST00000377815) was cloned from a human U2OS complementary DNA library using ATGGCGGGACGACACGATTGGCTAG and CTATGGAC TGCTGCATTGAAAAGGGC primers and inserted in frame into pEGFP (Clontech; Supplementary Fig. 6f). Alternatively, Bora- Δ exon1 (corresponding to ENST00000390667) was amplified with ATGGGAGATGTCAAGGAATCA AAGATGC and CTATGGACTGCTGCATTGAAAAGGGC primers using plasmid DNA as a template. Silent mutations generating RNAi-non-targetable Bora were introduced by the Quick Change mutagenesis kit (Stratagene). The following drugs were used: BI2536 (100 nM, Boehringer Ingelheim Pharma), MLN8054 (1–3 μ M, Millennium Pharmaceuticals), ZM447439 (5 μ M, AstraZeneca), doxorubicin and caffeine (Sigma). Polyclonal anti-PLK1 antibody (Supplementary Fig. 9b) was generated by immunizing rabbits with the carboxy-terminal domain of human PLK1, and rabbit antiserum was affinity purified. Polyclonal anti-Bora antibody was a gift from E. Nigg. Phosphospecific affinity purified pT210-PLK1 antibody (Supplementary Fig. 9b), and an Alexa488-labelled form of the same antibody, were from BD Pharmingen, anti-phospho-S10-histone H3 and mouse monoclonal anti-Myc (clone 9E10) from Upstate, anti-phospho-T495-MYT1 from Abgent, polyclonal anti-aurora A and anti-phospho-threonine from Cell Signaling, anti- α -tubulin from Sigma, Dyomics-647-conjugated anti- γ -tubulin from Exbio, and secondary antibodies Alexa-488, Alexa-568 and Alexa-633 from Molecular Probes.

Cell culture and transfections. Human osteosarcoma U2OS cells were grown in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% FCS (Integro), 2 mM L-glutamine, 100 U ml⁻¹ penicillin, and 100 μ g ml⁻¹ streptomycin. U2OS-TR (U2TR) cells stably expressing the tetracycline-repressor were prepared by transfection of pcDNA6/TR plasmid (Invitrogen) followed by blasticidin (10 μ g ml⁻¹) treatment and clonal selection. Cell lines expressing various PLK1 mutants under the control of tetracycline-inducible promoter were generated by transfection of U2TR cells with linearized pcDNA4/TO plasmids (Invitrogen), and stable clones were selected by zeocin (400 μ g ml⁻¹) treatment followed by clonal selection. Stable clones were grown in media containing Tet system approved fetal bovine serum (Clontech). For induction of expression, cells were treated for indicated times with tetracycline (1 μ g ml⁻¹). Two independent clones for each PLK1 mutant were tested. Transient transfections of plasmid DNA and siRNAs were carried out using the standard calcium phosphate technique, and siRNAs were transfected using Hiperfect reagent (Qiagen).

Cell synchronization, DNA damage and recovery. Cell synchronization was performed as described¹. In brief, cells were first co-transfected with the indicated targeting vectors together with GFP-spectrin (15:1), and synchronized by thymidine (2.5 mM, 24 h) treatment. To induce DNA damage in G2, medium was first replaced to release cells from the thymidine block, and at 5 h into this release cells were treated with doxorubicin (0.5 μ M) for 1 h. Cells were subsequently washed with PBS and incubated for 18 h in fresh medium supplemented with nocodazole to trap cells in mitosis. Where indicated, DNA-damage-signalling was silenced by addition of caffeine (5 mM). For reconstitution assays, expression of PLK1 and the respective mutants was induced by addition of tetracycline (1 μ g ml⁻¹). To determine the rate of recovery, cells were harvested 9 h after caffeine addition and fixed in ice-cold ethanol (70%). Cells were stained with anti-pS10-H3 and Cy5-conjugated anti-rabbit antibodies and counterstained with propidium-iodide. Cell cycle distribution was determined by flow cytometry counting 10⁴ events as described¹⁹.

Immunoprecipitations, and phosphorylation of PLK1 by aurora A. Cells were extracted in lysis buffer (50 mM HEPES, pH 7.4, 1 mM MgCl₂, 1 mM EGTA, 1% NP-40, 1 mM NaF, 1 mM Na₃VO₄, protease inhibitors), normalized for total protein content and incubated overnight (15 h) at 4 °C with polyclonal anti-PLK1 antibody immobilized on protein A (BioRad) or with S-protein agarose beads (Novagen). Immunocomplexes were extensively washed and analysed by immunoblotting. To analyse direct phosphorylation of PLK1 by aurora A, kinase-dead mutants of PLK1 were expressed as a maltose-binding protein fusion in *E. coli*. For co-expression, a plasmid for expressing aurora A was co-transformed with the PLK1 plasmid. The PLK1 proteins were batch-purified with amylose-agarose beads (New England Biolabs) and analysed by western blotting with PLK1-NT antibody (Upstate) and the pT210 antibody. The *in vitro* PLK1 phosphorylation assays shown in Fig. 3b were performed using GST-aurora A and kinase-dead MBP-PLK1 fusion proteins expressed and purified from *E. coli*. For each 53 μ l reaction, ~2 μ g of MBP-tagged kinase-dead PLK1 was incubated alone, or together with 0.5 μ g wild-type GST-aurora A. Kinase assays were performed at room temperature (22 °C), and 10- μ l aliquots taken at the indicated time points were analysed by western blotting with PLK1-NT and pT210 antibodies. The experiment shown in Fig. 3c was performed with 0.5 μ g GST-aurora A (wild-type or kinase-dead) and kinase-dead His-tagged PLK1 expressed and purified from *E. coli*. Alternatively, histone H3 (Roche) was used as a substrate. Kinase reactions were performed in aurora A kinase buffer (50 mM Tris, pH 7.4, 15 mM MgCl₂, 2 mM EGTA, 1 mM DTT, 0.5 mM Na₃VO₄, 100 μ M ATP and 5 μ Ci ³²P- γ -ATP). Where indicated, 0.2 μ g of GST-Bora, purified from *E. coli*, was added into the reaction mix. Kinase assays were analysed by autoradiography and western blotting with pT210 antibodies. For *in vitro* PLK1 phosphorylation using immunoprecipitated aurora A from mammalian cells, HEK293 cells expressing wild-type or kinase-dead Myc-tagged aurora A²⁷ were extracted in lysis buffer, and aurora A was immunoprecipitated using anti-Myc antibodies (clone 9E10, Upstate) immobilized on protein A/G beads (Biorad) at 4 °C for 3–4 h. Immunocomplexes were extensively washed in lysis buffer, followed by two washes in aurora kinase buffer. Kinase reactions were performed as mentioned above. Where indicated, 0.2 μ g of GST-Bora, purified from *E. coli*, was added into the reaction mix. Kinase assays were analysed by autoradiography and western blotting with pT210 antibodies.

Immunofluorescence and FRET probe. Fixation and antibody staining for immunofluorescence were performed as described³². Images show maximum intensity projections of deconvolved or confocal Z-stacks, acquired on a Zeiss LSM510 META or a Deltavision RT imaging system using NA 1.4 objectives. The FRET-based probe for monitoring PLK1 activity has been described previously¹⁵ and the Ala-mutant was generated by site-directed point mutagenesis (Stratagene). The CFP/YFP emission ratio after CFP excitation of U2OS cells, stably or transiently expressing the probe, was monitored on a Deltavision RT imaging system, using a NA 1.35 \times 40 objective. Control cells were imaged simultaneously by use of a two-well dish. Images were acquired every 20 min. The images were processed with ImageJ using the Ratio Plus plug-in (<http://rsb.info.nih.gov/ij/>). Quantification of immunofluorescence was performed as described³², measuring the centrosomal maximum intensity using a \times 100 NA 1.4 objective. To control for the specificity of the staining, pS-PLK1 transfected cells were seeded onto cells stably expressing YFP-tubulin, enabling the quantification of untreated and RNAi cells on the same coverslip.

28. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
29. Katayama, H., Zhou, H., Li, Q., Tatsuka, M. & Sen, S. Interaction and feedback regulation between STK15/BTAK/aurora-A kinase and protein phosphatase 1 through mitotic cell division cycle. *J. Biol. Chem.* **276**, 46219–46224 (2001).
30. Cheeseman, I. M. & Desai, A. A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Sci. STKE* **2005**, pl1 (2005).
31. Hoar, K. et al. MLN8054, a small-molecule inhibitor of Aurora A, causes spindle pole and chromosome congression defects leading to aneuploidy. *Mol. Cell. Biol.* **27**, 4513–4525 (2007).
32. Lindqvist, A., van Zon, W., Karlsson Rosenthal, C. & Wolthuis, R. M. Cyclin B1-Cdk1 activation continues after centrosome separation to control mitotic progression. *PLoS Biol.* **5**, e123 (2007).

LETTERS

Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B

Holger Rehmann¹, Ernesto Arias-Palomo², Michael A. Hadders³, Frank Schwede⁴, Oscar Llorca² & Johannes L. Bos¹

Epac proteins are activated by binding of the second messenger cAMP and then act as guanine nucleotide exchange factors for Rap proteins^{1,2}. The Epac proteins are involved in the regulation of cell adhesion³ and insulin secretion⁴. Here we have determined the structure of Epac2 in complex with a cAMP analogue (Sp-cAMPS) and RAP1B by X-ray crystallography and single particle electron microscopy. The structure represents the cAMP activated state of the Epac2 protein with the RAP1B protein trapped in the course of the exchange reaction. Comparison with the inactive conformation reveals that cAMP binding causes conformational changes that allow the cyclic nucleotide binding domain to swing from a position blocking the Rap binding site towards a docking site at the Ras exchange motif domain.

Previously we have determined the structure of Epac2 in the inactive state, showing that the access of Rap to the catalytic site, a helical hairpin in the CDC25-homology domain, is sterically blocked by the cyclic nucleotide binding (CNB) domains⁵. To understand the mechanism of activation and guanine nucleotide exchange factor (GEF) activity of Epac proteins, we solved the crystal structure of a deletion mutant of Epac2 lacking the first 305 amino acids in complex with RAP1B and Sp-cAMPS to 2.2 Å resolution (Fig. 1, Supplementary Table 1). Epac2Δ305 lacks the first CNB domain and the DEP domain, but, due to the presence of the critical second CNB, its activity is completely cAMP dependent⁵. Sp-cAMPS is a hydrolysis resistant cAMP analogue, which activates Epac normally⁶.

Comparison of Epac2Δ305•Sp-cAMPS•RAP1B with the structure of inactive full length Epac2 shows that the second CNB domain moves by 45 Å as a rigid body from one side of the helical hairpin, where it blocks the access of Rap to the catalytic site, to the other (Fig. 1b, c). To prove that this movement reflects the conformational changes that occur in the full-length protein, inactive full-length Epac2 and the active complex of full-length Epac2 and RAP1B were subjected to single particle electron microscopy (EM). A good match was obtained between the crystal structure⁵ and EM reconstructions of inactive Epac2 (Supplementary Fig. 1). Similarly, the X-ray structure of Epac2Δ305•Sp-cAMPS•RAP1B could be fitted in the EM density of Epac2•cAMP•RAP1B, whereby a continuous volume of extra density for the missing domains was obtained (Fig. 1d). Although the obtained resolution does not allow discrimination in the position of the missing domains, geometric constraint suggests the DEP domain to be in close proximity to the second CNB domain.

The function of the first CNB domain was investigated by mutational analysis: we find that this domain is not required to maintain the auto-inhibited state, and that cAMP binding to the first CNB domain is not required for activation (Supplementary Fig. 2). However, the first and the second CNB domain are arranged face to face to each other, so each is blocking the other's cAMP binding

site. Activation of Epac2 requires the displacement of the first CNB domain by stochastic opening of the regulatory region, which allows cAMP to access the second CNB domain and to induce activation of Epac2 (Supplementary Fig. 2). This is in agreement with the EM data, which show that the relative domain organization within the regulatory region changes upon cAMP binding.

The movement of the second CNB domain originates in the hinge, which connects the two carboxy-terminal β-strands of the second CNB domain to the central β-core of the domain (Fig. 1b). The movement can be best analysed by keeping the core of the CNB domain fixed and considering the catalytic region to move (Supplementary Movie 1). Two major effects are observed: (1) the hinge helix (and the whole catalytic region) swings closer to the core of the CNB domain; and (2) the last two turns of the hinge helix 'melt', resulting in a prolonged loop between the hinge helix and the C-terminal β-strands. This allows the C-terminal β-strands (and the whole catalytic region) to rotate about 90° sideways and to translate closer to the cAMP binding site (Fig. 2a).

By forming a rigid β-sheet-like structure with the first β-strand of the Ras exchange motif (REM) domain and the tip of the helical hairpin, the C-terminal β-strands anchor the CNB domain to the catalytic region. Even though structurally unaffected by cAMP binding, the C-terminal β-strands together with the first helix of the REM domain complete the cAMP binding site. The base of the cyclic nucleotide interacts with both elements, which are referred to as the lid. Several of these interactions are crucial for proper activation. For instance, Leu449 is packed in parallel to the adenine base (Fig. 2b). The maximal activity (k_{\max}), which is a measure of the efficiency with which cAMP shifts the equilibrium to the active site, is reduced in Epac2 L449A to less than 10% (Supplementary Fig. 3). The amino group of the adenine base interacts with Lys450 while cPuMP, which only differs from cAMP by the absence of this group, is a very weak activator of Epac2. The crucial need for the amino group thus explains the selectivity of Epac2 for cAMP over cGMP (Supplementary Fig. 4).

In addition to the interactions with Sp-cAMPS, the lid forms new contacts with the central β-core of the CNB domain (Fig. 2c). These interactions are clustered around Lys405, which is part of the phosphate binding cassette (PBC). The PBC is a highly conserved part of the central β-core and interacts with the phosphate-sugar moiety of cAMP⁷. At the corresponding position of Lys405, a totally conserved glutamic acid is found in all CNB domains except for those in Epac. The glutamic acid of non-Epac proteins crucially interacts with the 2'-OH group of the cyclic nucleotide and, in addition, with the lid structure of this proteins^{8–12}. However, no direct interaction is observed between Lys405 and Sp-cAMPS. Lys405 interacts with Tyr480 at the beginning of the REM domain and with the backbone

¹Department of Physiological Chemistry, Centre for Biomedical Genetics and Cancer Genomics Centre, University Medical Center, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands. ²Centro de Investigaciones Biológicas (CIB), Spanish National Research Council (CSIC), Ramiro de Maeztu 9, 28040 Madrid, Spain. ³Department of Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands. ⁴BIOLOG Life Science Institute, Flughafendamm 9a, 28199 Bremen, Germany.

oxygens of Asn 445 and Glu 445, which are part of the loop between the hinge helix and the amino-terminal β -strands. These interactions stabilize the active conformation of the lid, since Epac2 Y480F shows a k_{\max} which is reduced by 50% (data not shown).

In the active conformation, the β -core of the second CNB domain is packed against the REM domain (Fig. 1b). The emerging intra-molecular interface is dominated by van der Waals interactions clustered around a hydrogen bond between Gln 369 in the CNB domain and Tyr 551 in the REM domain. The k_{\max} is reduced in Epac2 Y551F and Epac2 Y551A to 70% of the wild-type protein (Supplementary Figs 3, 5).

In summary, there are three elements stabilizing the active conformation of Epac2: (1) the interaction of the cyclic nucleotide with the lid; (2) newly formed interactions of the central β -core and the lid; and (3) interactions between the core of the CNB domain and the REM domain. The REM domain is involved in all three of these elements and disrupting mutations decreased the maximal activity of Epac2, indicative of the importance of the proper stabilization of the active conformation.

Apart from stabilizing the active conformation, how is cAMP causing the transition to the active conformation? In the absence of cAMP, the PBC sterically prevents the hinge-helix from swinging

towards the central β -core. In agreement with a previously described model¹³, cAMP binding induces a tightening of the PBC thereby releasing this steric block (Supplementary Fig. 6). Once released, the hinge can undergo the aforementioned rearrangements, and the CNB domain can be trapped in the active conformation. This process crucially depends on the lid. Its β -sheet structure is unique to the CNB domains of Epac, since helical structures are found in PKA^{9,10} and cyclic nucleotide regulated ion channels^{11,12} (Supplementary Fig. 7).

How is Epac interacting with its substrate RAP1B? Binding of RAP1B to the CDC25-homology domain induces an upwards bending of the helical hairpin towards the REM domain, which corresponds to a C_{α} -backbone displacement of at most 2 Å at its tip. A similar effect was observed in the complex of HRAS and its GEF Sos, which was attributed to a repositioning of the REM domain^{14,15}. The REM domain of the active Epac2 is slightly affected by its interaction with the CNB domain. However, these changes are different and

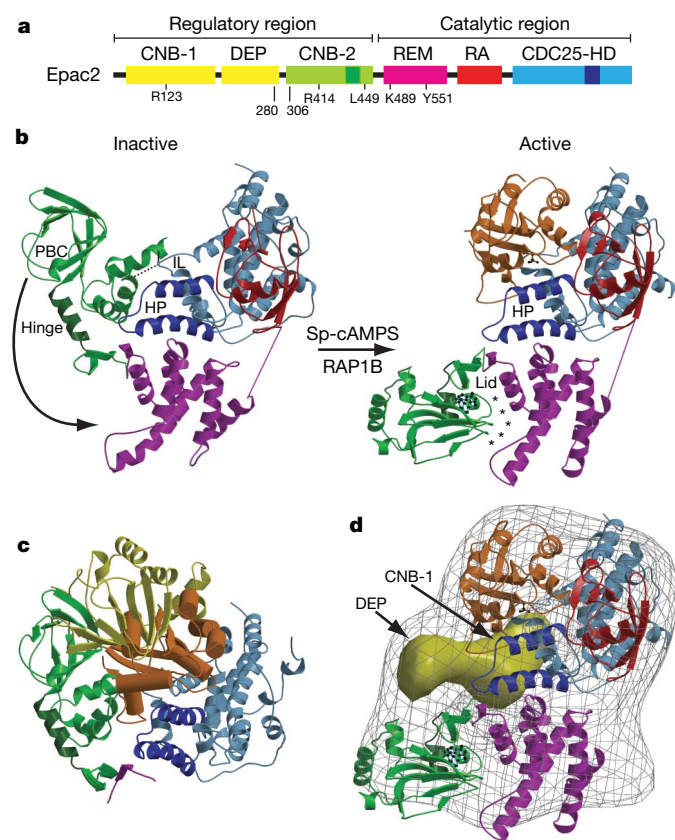


Figure 1 | Active Epac 2. **a**, Domain organization of Epac2. Residues that were subjected to mutational analysis are indicated. The same colour code is used throughout the figures. Hinge (residues 432–445, dark green); helical hairpin (residues 906 to 946, dark blue). CDC25-HD, CDC25 homology domain; CNB, cyclic nucleotide binding domain; DEP, Dishevelled, Egl-10, Pleckstrin domain; RA, Ras-association domain; REM, Ras-exchange motif. **b**, Left, inactive Epac2 (first CNB and DEP domain omitted); right, active Epac2Δ305•Sp-cAMPS•RAP1B. RAP1B is shown orange; Sp-cAMPS and SO_4^{2-} are shown in ball and stick representation. Arrow, movement of the second CNB domain; straight lines, missing connectivity; dotted lines, ionic latch (IL); asterisks, interface between the REM and the CNB domain; HP, helical hairpin; PBC, phosphate binding cassette. **c**, RAP1B placed into the inactive structure. **d**, The crystal structure of Epac2Δ305•Sp-cAMPS•RAP1B was fitted into the EM density reconstruction (grey grid) of full length Epac2•cAMP•RAP1B. Yellow surface, difference density.

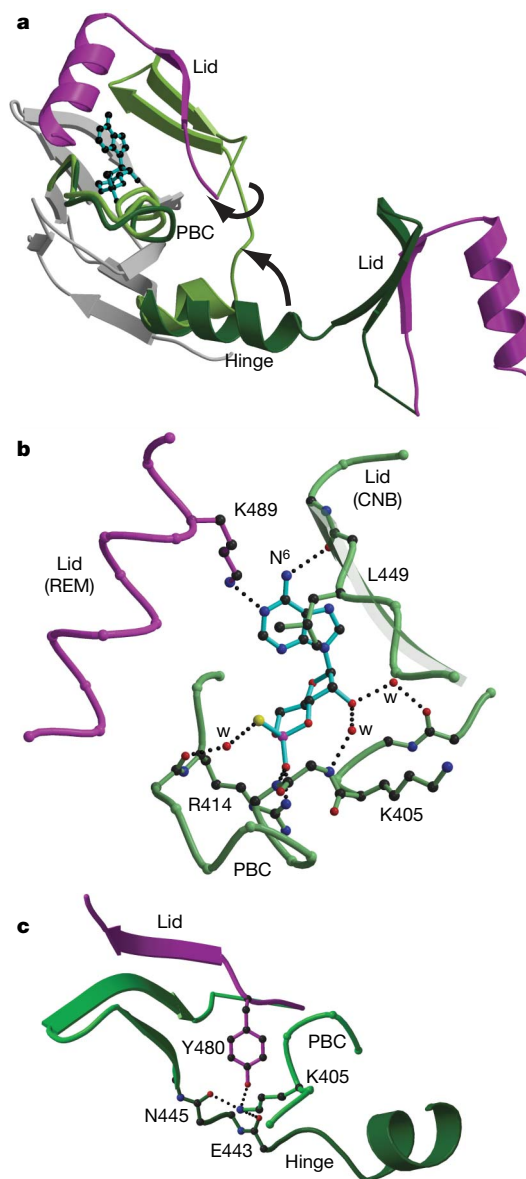


Figure 2 | Sp-cAMPS induced conformational changes. **a**, Superposition of the active and inactive second CNB domain. The arrows indicate the movement of the hinge and the lid region. Light green, active conformation; dark green, inactive conformation; grey, no difference in conformation. **b**, Interactions of Sp-cAMPS with the CNB domain and the REM domain. Hydrogen bonds are shown by dotted lines; w, water. **c**, Interaction of Lys 405 with the hinge-lid region.

much more subtle than in Sos and might be attributed to the constraints of crystal packing.

The nucleotide binding site of G proteins consists mainly of switch I and II, the P-loop that amply interacts with the phosphate moiety of the nucleotide, and the NKxD motif that interacts with the base.

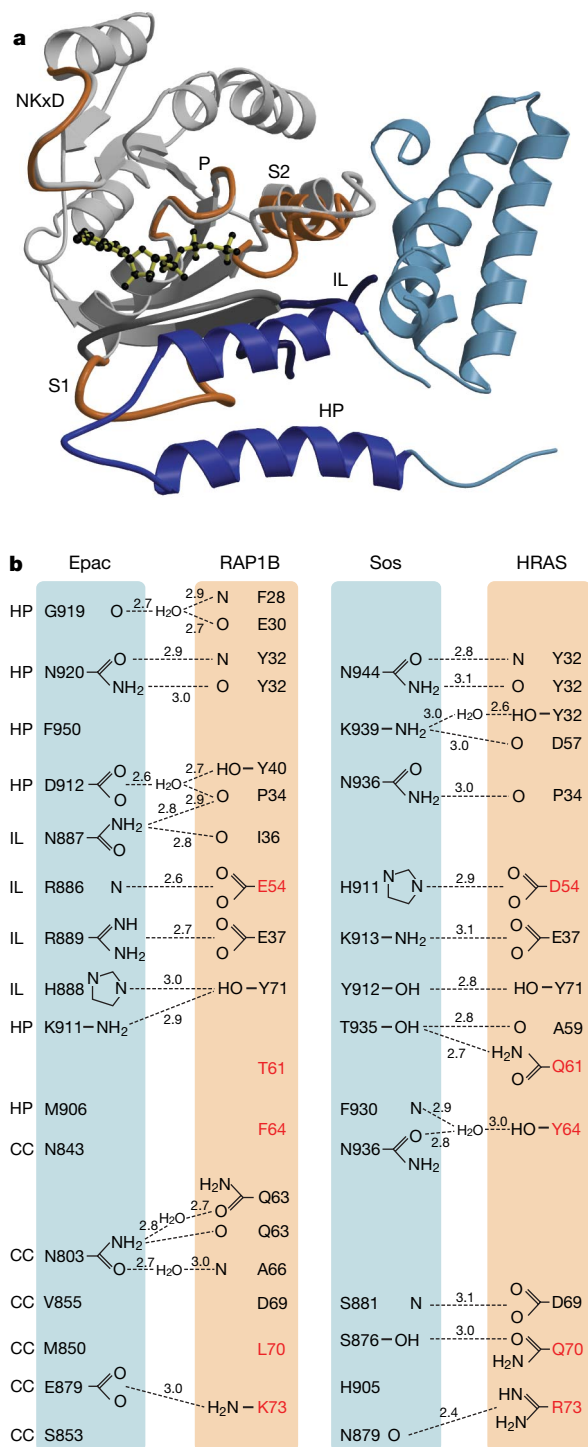


Figure 3 | Interactions between Epac2 and RAP1B. **a**, The structure of RAP1A (PDB entry 1c1y; grey) bound to GTP (ball-and-stick) was superimposed onto the nucleotide free RAP1B in Epac2Δ305•Sp-cAMPS•RAP1B, which is shown in orange only where it adopts a different conformation. HP, helical hairpin (blue); IL, ionic latch loop (dark blue); NKxD, NKxD motif; P, P-loop; S1, switch 1; S2, switch 2. **b**, Comparison of interaction in the Epac2Δ305•Sp-cAMPS•RAP1B and the Sos•HRAS complex. Residues not conserved in RAP1B and HRAS are highlighted in red. Hydrogen bonds are indicated by dotted lines, and distances are given in Å. CC, central core.

Epac2 inserts the helical hairpin of the CDC25-homology domain into the nucleotide binding site of Rap, thereby bending switch I and opening the nucleotide binding site (Fig. 3a). Switch II becomes deformed, and is thus disabled from interacting properly with the nucleotide by being tightly packed against the core of the CDC25-homology domain (Fig. 3a, Supplementary Fig. 8). In contrast, the conformation of neither the P-loop nor the NKxD motif is changed by Epac2 (Fig. 3a). The P-loop is stabilized by a SO_4^{2-} anion, which takes the position of the β -phosphate and mimics its interactions (Supplementary Fig. 8). The absence of the SO_4^{2-} anion would probably result in a collapse of the P-loop, as seen in the HRAS•Sos complex¹⁶. Nevertheless, the direct action of CDC25-homology domains is limited to the switch regions as shown here.

Three major elements of the CDC25-homology domain contribute to the interaction with RAP1B: the helical hairpin, the part of the domain core to which the switch II region is packed, and the ionic latch loop. In the inactive conformation, residues in the ionic latch loop interact with the N terminus of the second CNB domain. The catalytic activity of the ionic latch loop mutant Epac2 R886A was found to be drastically reduced⁵, demonstrating that in the inactive conformation, residues crucial for catalysis are directly masked by an interaction with the second CNB domain. In the Epac2Δ305•Sp-cAMPS•RAP1B complex, Arg 886^{Epac2} and Arg 889^{Epac2} are arranged in parallel with Glu 37^{RAP1B}, and Arg 889^{Epac2} and Glu 37^{RAP1B} are in hydrogen bonding distance (Fig. 3b). Indeed, the ability of Epac2 to activate RAP1B E37A is reduced (Supplementary Fig. 8).

Even though the general interaction mode is conserved between Epac2Δ305•Sp-cAMPS•RAP1B and Sos•HRAS, there are remarkable differences in detail (Fig. 3b). HRAS and RAP1B are characterized by high sequence homology, and selectivity has thus to be determined by the limited number of non-conserved residues. However, attempts to change selectivity by single point mutations in HRAS have been unsuccessful (so far; data not shown), suggesting that selectivity emerges from the whole assembly of non-conserved residues. Sos and Epac2 interact with most of the conserved residues in RAP1B and HRAS differently and the ionic latch loop is even two residues longer in Epac than in Sos (Fig. 3b). This suggests that the catalytic mechanism is conserved but not the detailed interaction, which allows the establishment of selectivity.

The structural characterization of the activated state of Epac2 demonstrates how the small ligand cAMP induces huge conformational changes within the protein. The incoming cAMP molecule causes small changes in the immediate binding site, which result in the release of steric constraints and finally the collapse of the stable auto-inhibited conformation. The obtained structural flexibility allows 'undirected' rearrangements, whereby Epac is finally trapped in the active conformation by the establishment of the full cAMP binding site (Supplementary Movie 1).

METHODS SUMMARY

Epac2 proteins were purified from *Escherichia coli* using a glutathione S-transferase (GST) tag and gel filtration. The GST tag was removed by thrombin cleavage. A C-terminal truncated version of RAP1B was used. Several deletion constructs of Epac2 were tested for their ability to form crystals. A protein solution of Epac2Δ305, RAP1B and Sp-cAMPS was prepared by mixing 8.7 g l⁻¹ Epac2Δ305, 3.4 g l⁻¹ RAP1B, and 3 mM Sp-cAMPS in buffer containing 40 mM TrisHCl pH 7.5, 50 mM NaCl, 2.5% glycerol and 1 mM EDTA. Crystals were grown at 277 K in sitting drops using a reservoir solution containing 0.4 M $(\text{NH}_4)_2\text{SO}_4$, 1.2 M LiSO_4 and 0.1 M sodium citrate pH 5.6. Crystals diffract to 2.2 Å resolution. The structure was solved by molecular replacement.

The samples for EM were prepared under similar conditions, applied to carbon-coated grids and negatively stained with 2% uranyl formate. As a control, the inactive Epac2 was used. The obtained density reconstruction corresponds to a resolution of 23 Å and 30 Å for Epac2•RAP1B and Epac2, respectively.

For the GEF assays, RAP1B was loaded with the fluorescent GDP analogue 2'-3'-O-(N-methylanthraniloyl)-guanosine diphosphate (mGDP). The fluorescence intensity of RAP1B bound mGDP is approximately twice that of free mGDP, and thus nucleotide exchange can be observed as decay in fluorescence

upon addition of an excess of unlabelled GDP. The speed of the decay depends on Epac2 activity and thus on the cyclic nucleotide concentration.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 February; accepted 20 June 2008.

Published online 27 July 2008.

- de Rooij, J. *et al.* Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* **396**, 474–477 (1998).
- Kawasaki, H. *et al.* A family of cAMP-binding proteins that directly activate Rap1. *Science* **282**, 2275–2279 (1998).
- Bos, J. L. Linking Rap to cell adhesion. *Curr. Opin. Cell Biol.* **17**, 123–128 (2005).
- Seino, S. & Shibasaki, T. PKA-dependent and PKA-independent pathways for cAMP-regulated exocytosis. *Physiol. Rev.* **85**, 1303–1342 (2005).
- Rehmann, H., Das, J., Knipscheer, P., Wittinghofer, A. & Bos, J. L. Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature* **439**, 625–628 (2006).
- Rehmann, H., Schwede, F., Døskeland, S. O., Wittinghofer, A. & Bos, J. L. Ligand-mediated activation of the cAMP-responsive guanine nucleotide exchange factor Epac. *J. Biol. Chem.* **278**, 38548–38556 (2003).
- Canaves, J. M. & Taylor, S. S. Classification and phylogenetic analysis of the cAMP-dependent protein kinase regulatory subunit family. *J. Mol. Evol.* **54**, 17–19 (2002).
- Yagura, T. S. & Miller, J. P. Mapping adenosine cyclic 3',5'-phosphate binding sites on type I and type II adenosine cyclic 3',5'-phosphate dependent protein kinases using ribose ring and cyclic phosphate ring analogues of adenosine cyclic 3',5'-phosphate. *Biochemistry (Mosc.)* **20**, 879–887 (1981).
- Su, Y. *et al.* Regulatory subunit of protein kinase A: Structure of deletion mutant with cAMP binding domains. *Science* **269**, 807–813 (1995).
- Diller, T. C., Madhusudan, Xuong N. H. & Taylor, S. S. Molecular basis for regulatory subunit diversity in cAMP-dependent protein kinase: Crystal structure of the type II beta regulatory subunit. *Structure (Camb.)* **9**, 73–82 (2001).
- Clayton, G. M., Silverman, W. R., Heginbotham, L. & Morais-Cabral, J. H. Structural basis of ligand activation in a cyclic nucleotide regulated potassium channel. *Cell* **119**, 615–627 (2004).
- Zagotta, W. N. *et al.* Structural basis for modulation and agonist specificity of HCN pacemaker channels. *Nature* **425**, 200–205 (2003).
- Rehmann, H., Wittinghofer, A. & Bos, J. L. Capturing cyclic nucleotides in action: Snapshots from crystallographic studies. *Nature Rev. Mol. Cell Biol.* **8**, 63–73 (2007).
- Margarit, S. M. *et al.* Structural evidence for feedback activation by Ras•GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* **112**, 685–695 (2003).
- Freedman, T. S. *et al.* A Ras-induced conformational switch in the Ras activator Son of sevenless. *Proc. Natl Acad. Sci. USA* **103**, 16692–16697 (2006).
- Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, J. The structural basis of the activation of Ras by Sos. *Nature* **394**, 337–343 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Gros for access to crystallization robots, M. Rensen-De Leeuw for technical assistance, A. Wittinghofer for reading the manuscript, the European Synchrotron Radiation Facility for providing synchrotron facilities, the scientists at ID23-1 for help with data collection, and the Galicia Supercomputer Centre (CESGA) and the Barcelona Supercomputing Centre for computing resources. H.R. is a recipient of the Hendrik Casimir-Karl Ziegler-Forschungspreis of the Nordrhein-Westfälischen Akademie der Wissenschaften and the Koninklijke Nederlandse Akademie van Wetenschappen. E.A.-P. and O.L. are supported by the Autonomous Region of Madrid, F.S. by the Bremer Innovationsagentur, O.L. by the Spanish Ministry of Education and Science (MEC) and the Red Temática de Investigación cooperativa en Cáncer (RTICC), and J.L.B. by the Chemical Sciences and the Netherlands Genomic Initiative of the Netherlands Organisation for Scientific Research (NWO).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Co-ordinates have been deposited with the Protein Data Bank under accession number 3cf6. The EM map of activated full length Epac2 bound to RAP1B has been deposited in the 3D EM database (<http://www.ebi.ac.uk/msd/>) under accession code EMD-1510. Correspondence and requests for materials should be addressed to H.R. (h.rehmann@UMCcutrecht.nl).

METHODS

Protein preparation. All Epac2 proteins (Epac2 (also known as RapGEF 4) *Mus musculus*, accession number AF115480) were expressed as GST-fusion (pGEX4T, Pharmacia) in the *E. coli* strain CK600K similar to the procedure described for Rap1B (ref. 17). RAP1B (RAP1B, *Homo sapiens*, accession number X08004, amino acids 1–167) was purified as described¹⁸.

Crystallography. The following constructs were screened for their ability to form crystals in complex with RAP1B and cyclic nucleotide: Epac2 1–993 (Epac2 full length), Epac2 280–993 (Epac2Δ279), Epac2 288–293 (Epac2Δ287), Epac2 293–993 (Epac2Δ292), Epac2 306–993 (Epac2Δ305) and Epac2 319–993 (Epac2Δ318). Only Epac2Δ305 was found to form crystals. The activity of Epac2Δ305 is, as the full length protein, dependent on cAMP. However, it was previously shown that in comparison to full length Epac2 or Epac2Δ279, cAMP binding to Epac2Δ305 shifts the equilibrium between the inactive and the active conformation more efficiently to the active state⁵. This effect was attributed to a partial destabilization of interactions formed by the ionic latch, which stabilizes the inactive conformation⁵. Thus the complex formed by Epac2Δ305 might be more homogeneous in comparison to other Epac2 variants and might crystallize more easily. In addition, a longer N terminus might be incompatible with the obtained crystal matrix.

The complex of Epac2Δ305 with RAP1B and Sp-cAMPS (6-(6-amino-purin-9-yl)-2-thioxo-tetrahydro-2-furo[3,2-d][1,3,2]dioxaphosphinine-2,7-diol) was prepared by mixing 8.7 g l⁻¹ Epac2Δ305, 3.4 g l⁻¹ RAP1B, and 3 mM Sp-cAMPS in buffer containing 40 mM TrisHCl pH 7.5, 50 mM NaCl, 2.5% glycerol and 1 mM EDTA. Crystals of the complex were grown at 277 K in sitting drops using a reservoir solution containing 0.4 M (NH₄)₂SO₄, 1.2 M LiSO₄ and 0.1 M sodium citrate pH 5.6. Crystals were cryoprotected in a solution containing the mother liquor supplemented with 2.1 M LiSO₄ and flash frozen in liquid nitrogen. Data were collected at 100 K at beamline ID23-1 of ESRF. Crystals diffract to 2.2 Å resolution and belong to the space group I212121 with one molecule per asymmetric unit (Supplementary Table 1).

The data were processed with XDS¹⁹. Molecular replacement was carried out in MOLREP²⁰. In a first step, the catalytic region (residues 482–991) of the auto-inhibited Epac2 (PDB code 2BYV) was used as a search model. The solution was fixed and in a second step, HRAS from HRAS•Sos (PDB code 1BKD) was used as a poly-alanine model for RAP1B. In a third step, the second CNB domain (residues 320–432) of the auto-inhibited Epac2 were added. The program O²¹ was used to build the model into 2F_o – F_c and F_o – F_c maps and refinement was carried out with REFMAC²². Residues 306–309, 463–477, 613–642, 953–961 and 991–993 of Epac2 and 1–2, 45–49 and 135–141 of RAP1B are not visible. The Ramachandran plot depicts 93.6% of main chain torsion angles in the most favoured regions, 6.1% in additional allowed regions, with 0 residues in disallowed regions.

Two molecules of Sp-cAMPS were unambiguously identified in the electron density. One is bound to the CNB domain and the second one is sandwiched in a crystal contact interface between parts of a CNB domain and parts of two REM domains (Supplementary Fig. 8). The second Sp-cAMPS molecule is attributed to the crystallization conditions, as it interacts with residues that are not conserved within Epac2 proteins and any biochemical indication for cAMP binding outside the canonical binding site is lacking.

Figures were generated using the programs Molscript²³, Bragi²⁴ and Raster3D²⁵.

Electron microscopy and 3D reconstruction. The complex of full length Epac2 was prepared by incubating Epac2 with GST-RAP1B and cAMP and purified to homogeneity by the GST, which was then cleaved off. Observations were performed in a JEOL 1230 transmission electron microscope operating at 100 kV and a FEI Tecnai G² operated at 200 kV. Micrographs were recorded at a magnification of 50,000 under low dose conditions and digitized using a Minolta Dimage Scan Multi Pro scanner at 2,400 d.p.i.; 8,712 and 7,412 images of Epac2•RAP1B•cAMP and of Epac2, respectively, were extracted using the program Boxer²⁶. The contrast transfer function of the microscope for each micrograph was estimated using CTFFIND3 (ref. 27) and corrected using Bsoft²⁸. The extracted particles were averaged to a final 4.2 Å per pixel and subjected to iterative refinement using a combined approach with EMAN²⁶ and 3D maximum likelihood classification (ML3D)²⁹. Initial 2D reference-free classification of the data was performed using EMAN and 2D maximum-likelihood methods implemented in XMIPP^{30,31}. Several initial template volumes for refinement were built from selected 2D averages through common-lines or constructed as gaussian blobs with the rough dimensions of the protein. The crystallographic struc-

tures were never used as input in the refinement to avoid any model bias. Models converged after refinement with EMAN into a 3D reconstruction that was the input of ML3D²⁹ to classify the data into more homogeneous groups. For this purpose, three initial template volumes were prepared by splitting the data into three random subsets. Starting from these seeds, ML3D classification converged within 20 cycles and the volume accounting for the largest number of particles was selected as the most representative. The resolution of the final 3D maps was estimated using two independent reconstructions obtained using even and odd halves of the particle data set (*eotest* command in EMAN²⁶), which corresponded to 30 Å and 23 Å using a 0.5 correlation coefficient for Epac2 and Epac2•RAP1B, respectively. The final volumes were low-pass filtered accordingly and visualized using UCSF Chimera³².

Fitting of the atomic structures into the EM densities was performed using COLORES as implemented in SITUS³³, and the ADP_EM platform³⁴, testing both possible hands of the final reconstructions. Both programs provided a better correlation coefficient (0.79 and 0.78) for one of the hands. A difference map between the crystallographic structure of truncated Epac2Δ305•Sp-cAMPS•RAP1B complex fitted into the EM map and the experimental EM reconstruction of Epac2•cAMP•RAP1B was calculated using SPIDER³⁵ after filtering the atomic coordinates at the resolution of the EM structure.

Biochemistry. GEF assays were performed as described previously¹⁷. For mutational analysis, Epac2Δ279 was chosen when the first CNB domain was not object of the investigation. Epac2Δ279 can be obtained in much higher amounts than full length Epac2 and any complication by the first CNB domain can be excluded. It addition it was previously shown that in comparison to full length Epac2 or Epac2Δ279 cAMP binding to Epac2Δ305 shifts the equilibrium between the inactive and the active conformation more efficiently to the active state⁵.

- Rehmann, H. Characterization of the activation of the Rap-specific exchange factor Epac by cyclic nucleotides. *Methods Enzymol.* **407**, 159–173 (2006).
- Herrmann, C., Horn, G., Spaargaren, M. & Wittinghofer, A. Differential interaction of the ras family GTP-binding proteins H-Ras, Rap1A, and R-Ras with the putative effector molecules Raf kinase and Ral-guanine nucleotide exchange factor. *J. Biol. Chem.* **271**, 6794–6800 (1996).
- Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
- Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56**, 1622–1624 (2000).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Kraulis, P. J. Molscript — a program to produced both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Schomburg, D. & Reichelt, J. Bragi — a comprehensive protein modeling program system. *J. Mol. Graphics* **6**, 161–165 (1988).
- Merritt, E. A. & Murphy, M. E. P. Rasr3D version 2.0 — A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
- Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
- Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
- Heymann, J. B. & Belnap, D. M. Bsoft: Image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
- Scheres, S. H. et al. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. *Nature Methods* **4**, 27–29 (2007).
- Scheres, S. H. et al. Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* **348**, 139–149 (2005).
- Sorzano, C. O. et al. XMIPP: A new generation of an open-source image processing package for electron microscopy. *J. Struct. Biol.* **148**, 194–204 (2004).
- Pettersen, E. F. et al. UCSF Chimera — a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Wriggers, W., Milligan, R. A. & McCammon, J. A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* **125**, 185–195 (1999).
- Garzon, J. I., Kovacs, J., Abagyan, R. & Chacon, P. ADP_EM: Fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics* **23**, 427–433 (2007).
- Frank, J. et al. SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).

LETTERS

Direct observation of the mechanochemical coupling in myosin Va during processive movement

Takeshi Sakamoto¹, Martin R. Webb², Eva Forgacs³, Howard D. White³ & James R. Sellers¹

Myosin Va transports intracellular cargoes along actin filaments in cells¹. This processive, two-headed motor takes multiple 36-nm steps in which the two heads swing forward alternately towards the barbed end of actin driven by ATP hydrolysis². The ability of myosin Va to move processively is a function of its long lever arm, the high duty ratio of its kinetic cycle and the gating of the kinetics between the two heads such that ADP release from the lead head is greatly retarded^{3–10}. Mechanical studies at the multiple- and the single-molecule level suggest that there is tight coupling (that is, one ATP is hydrolysed per power stroke), but this has not been directly demonstrated^{4,5,11}. We therefore investigated the coordination between the ATPase mechanism of the two heads of myosin Va and directly visualized the binding and dissociation of single fluorescently labelled nucleotide molecules, while simultaneously observing the stepping motion of the fluorescently labelled myosin Va as it moved along an actin filament. Here we show that preferential ADP dissociation from the trail head of mouse myosin Va is followed by ATP binding and a synchronous 36-nm step. Even at low ATP concentrations, the myosin Va molecule retained at least one nucleotide (ADP in the lead head position) when moving. Thus, we directly demonstrate tight coupling between myosin Va movement and the binding and dissociation of nucleotide by simultaneously imaging with near nanometre precision.

The ability to visualize the binding of fluorescent nucleotides to myosin in the light microscope has been limited by technical problems, such as the nonspecific binding of the fluorescent nucleotide to the coverslip, low quantum yield and rapid photobleaching. This has limited the maximum nucleotide concentration that could be used with analogues such as Cy3-labelled ATP to less than 100 nM^{12–15}. To overcome these problems, we have used a fluorescent ATP analogue (3'-(7-diethylaminocoumarin-3-carboxylamino)-3'-deoxyadenosine-5'-triphosphate; deac-aminATP), in which the fluorescence emission increases ~25-fold (Supplementary Fig. 1) when bound to a heavy meromyosin-like fragment of myosin Va (MyoV-HMM) in solution^{16,17}. The kinetic mechanism of MyoV-HMM using deac-aminATP as a substrate has been thoroughly studied, including the extent of gating that occurs between the two heads during movement^{10,17}. In brief, deac-aminATP binds 3-fold faster to MyoV than ATP does, and deac-aminADP dissociates 10–20-fold slower than ADP¹⁷. When MyoV-HMM is bound to actin by both heads, the release rate of deac-aminADP from the lead head is decreased by about 30-fold compared to the unstrained rate¹⁰. The processive run length of MyoV-HMM on actin using deac-aminATP as a substrate is shorter ($1,050 \pm 80$ nm) than when using ATP alone ($1,950 \pm 160$ nm) (Supplementary Figure 2a). The maximal velocity of movement on actin at saturating deac-aminATP is 120 nm s^{-1} , approximately 8–10-fold less than observed with ATP¹⁰ (Supplementary Fig. 2b).

Deac-aminADP that was non-specifically bound to a coverslip surface in the absence of MyoV-HMM was visualized using an electron

multiplying charged coupled device (EMCCD) camera at a camera gain level of 1,000 (the scale for gain is 0–1,000; Fig. 1a, e). The gain on the camera chip was then reduced to 400, at which the intensity of the nonspecifically bound deac-aminADP spots was considerably reduced (Fig. 1b, f). However, at the same gain (400) and collection time of 330 ms, deac-aminADP that was bound to MyoV-HMM on the coverslip (Fig. 1c, g) had a sufficiently high intensity (>10,000 photons) to fit the point-spread function of a single spot and so determine its precise nanometre localization¹⁸ (Supplementary Fig. 3). At the single-molecule level, we found a 4-fold enhancement of the fluorescent intensity of deac-aminADP on binding to MyoV-HMM.

We exchanged Alexa-Fluor-568-labelled calmodulins for the endogenous calmodulin bound to the neck region of MyoV-HMM. On average, each calmodulin contained 1.8 Alexa Fluor 568 moieties, and three Alexa-Fluor-568-labelled calmodulins were exchanged per MyoV-HMM, making it much brighter than myosin fused to GFP molecules or containing a single Cy3- or rhodamine-labelled calmodulin that had been previously used for single-molecule studies^{3,18}. Similar estimates for labelling ratios were obtained by using spectrophotometric techniques in solution or by examining the photobleaching kinetics of the molecules in the microscope (Supplementary Fig. 4). This allowed the Alexa-Fluor-568–MyoV-HMM to be as bright as the deac-aminonucleotides, and permitted the same camera and camera settings to be used to image both (Fig. 1d, h and Supplementary Fig. 3).

We simultaneously visualized Alexa-Fluor-568–MyoV-HMM and deac-aminonucleotide during processive movement on actin filaments *in vitro* (Fig. 2a, b; Supplementary Fig. 5 and Supplementary Movie). The Alexa-Fluor-568–MyoV-HMM and the deac-aminonucleotide fluorescence moved in the same direction at the same rate and on the same actin filaments (Fig. 2a, b and Supplementary Figs 5 and 6). The fluorescent signal from Alexa-Fluor-568–MyoV-HMM moved in 36-nm steps as would be expected from a molecule, in which both heads were labelled (Fig. 2a), albeit there is the possibility of minor differences between the alternating step sizes due to unevenness in the labelling of the two heads (see Supplementary Fig. 7 for an example of 'limping' movement). The deac-aminonucleotide moved in 18-nm steps. One step occurred simultaneously with the MyoV-HMM step, whereas the other step occurred during a dwell in the MyoV-HMM movement (Fig. 2b). These observations from a single trace are reinforced by examining histograms of the MyoV-HMM step size (which shows a peak of 36 ± 7 nm; Fig. 2d) and of the deac-aminonucleotide step size (which shows two peaks of 18 ± 7 nm and 36 ± 9 nm; Fig. 2e). The larger, 36 nm values for the deac-aminonucleotide movement are expected to result when two 18-nm movements occurred without a discernable dwell between them. This is calculated to occur 22–37% of the time ($1 - e^{-kt}$) on the basis of the deac-aminonucleotide association

¹Laboratory of Molecular Physiology, National Heart, Lung and Blood Institute, Bethesda, Maryland 20892, USA. ²MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK. ³Department of Physiological Sciences, Eastern Virginia Medical School, Norfolk, Virginia 23507, USA.

and dissociation rate constants measured in Fig. 3 and the 330 ms data acquisition time. The intensity of the deac-aminonucleotide signal integrated from a 12×12 pixel ($840 \times 840 \text{ nm}^2$) area surrounding the molecule at each frame showed a bimodal distribution in which one peak contained a factor of two more photons per frame than the other (Supplementary Fig. 8). The photon count in the smaller peak represents one deac-aminonucleotide per MyoV-HMM, whereas that in the other represents two per MyoV-HMM. Note that this nucleotide has similar fluorescence intensity when bound as MyoV-HMM-ADP, MyoV-HMM-ADP- P_i or

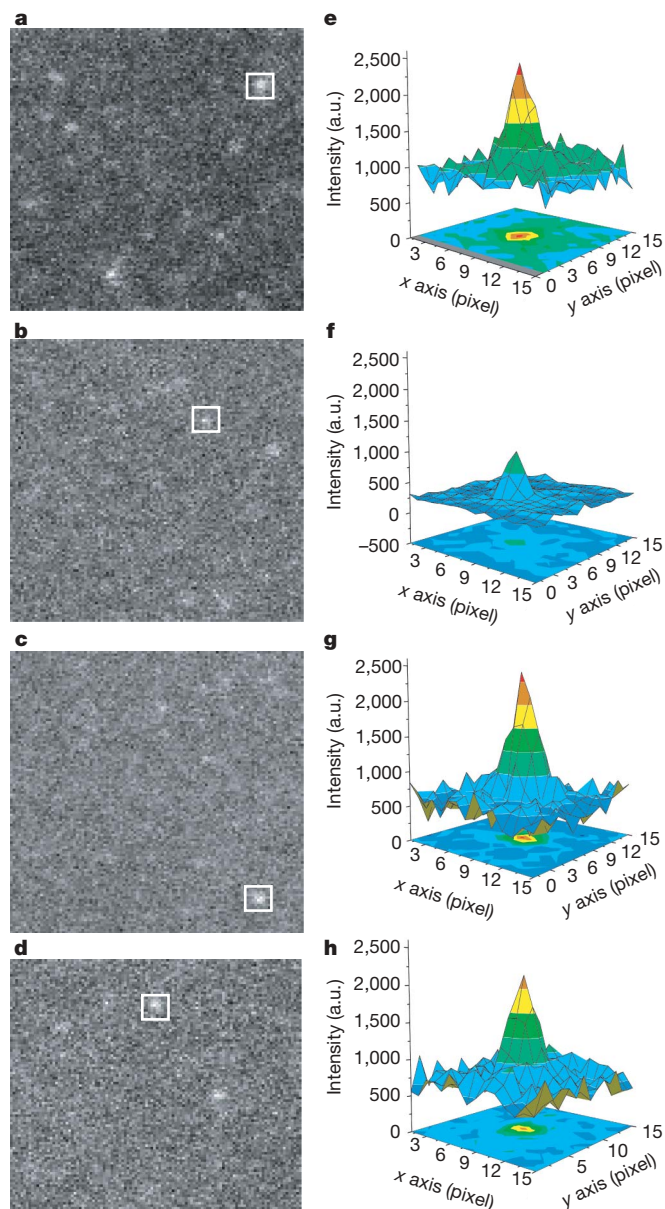


Figure 1 | Imaging deac-aminADP and Alexa-Fluor-568-MyoV-HMM. **a–d**, TIRF-microscopic images (110×100 pixels) are shown. Deac-aminADP was imaged at 442 nm (**a–c**); Alexa-Fluor-568-MyoV-HMM was imaged at 568 nm (**d**). Two-dimensional intensity profiles from each white square in **a–d** are shown in **e–h**. **a, e**, Deac-aminADP bound directly on the coverslip, with maximal camera gain (1,000). **b, f**, Deac-aminADP bound directly on the coverslip surface at a camera gain of 400. **c, g**, Deac-aminADP bound to MyoV-HMM at a camera gain of 400. **d, h**, Alexa-Fluor-568-MyoV-HMM bound to the surface at a camera gain of 400. All data were taken with an iXon+ camera (DV897, Andor technology) at 10 MHz readout at a constant laser power. The background level was fixed at about 750 (arbitrary units, a.u.) intensity. Scale bar, $2 \mu\text{m}$

MyoV-HMM-ATP, and thus, we cannot discriminate between different nucleotide states of a single head by intensity¹⁷. Using this criterion, the normalized intensity of the deac-aminonucleotide signal was also plotted as a function of time (Fig. 2c) and was shown to change from a value of one to two during each MyoV-HMM step and then decrease from a value of two to one during the MyoV-HMM dwell period.

The model to account for the 36-nm Alexa-Fluor-568-MyoV-HMM steps and the 18-nm deac-aminonucleotide steps is shown in Fig. 2f. Initially, MyoV-HMM has deac-aminADP bound to both heads and the position of the Alexa-Fluor-568-MyoV-HMM and the deac-aminonucleotide spots are coincident (step 1). Deac-aminADP is then released from the trail head, which results in the position of the deac-aminonucleotide signal advancing by 18 nm

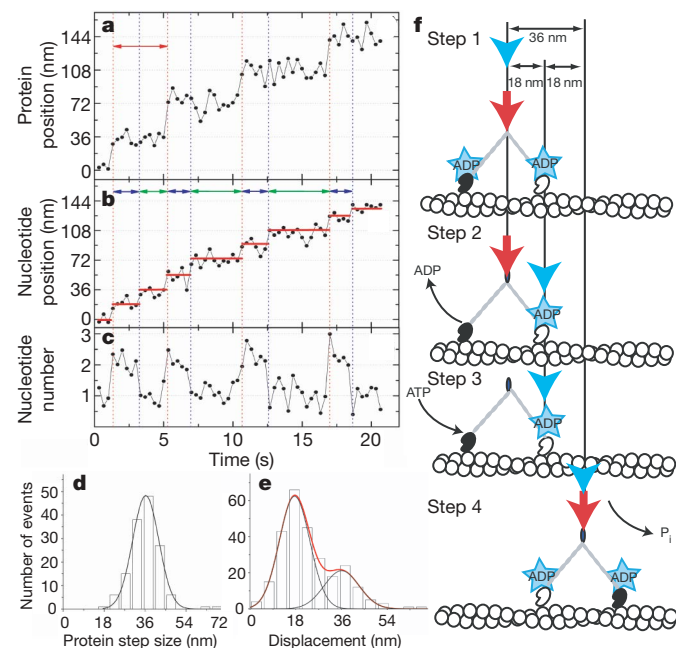


Figure 2 | Correlation between the movement of MyoV-HMM and the binding/dissociation of deac-aminonucleotide. Images of Alexa-Fluor-568-MyoV-HMM and deac-aminADP fluorescence were acquired simultaneously with a Dual-View system. The excitation/emission wavelengths of Alexa Fluor 568 are distinct from those of deac-aminADP, allowing simultaneous visualization of the MyoV-HMM and the nucleotide. The photons from the spots were acquired using 330 ms integrations and the point spread function from each spot was fit with a two-dimensional Gaussian to determine the location of the fluorophor(es) at each time point¹⁸. The deac-aminADP concentration was 200 nM. **a**, The protein fluorescence data show ~ 36 -nm steps, which are marked by red vertical dotted lines. Red double-headed arrows delineate the dwell time of a step. **b**, The deac-aminonucleotide stepping events are marked by alternating red and blue vertical dotted lines. Red vertical dotted lines are steps of both MyoV-HMM and deac-aminonucleotide, whereas blue vertical dotted lines show only stepping of deac-aminonucleotide. Individual spots move in a stepwise manner in the same direction as the MyoV-HMM. Dwell times are marked by blue and green double arrows. Red horizontal lines mark the average position of the spot during a pause. **c**, Normalized intensity of the deac-aminonucleotide fluorescence. **d**, Step-size histogram for the movement of Alexa-Fluor-568-MyoV-HMM ($n = 145$ steps, 38 MyoV-HMM molecules). The curve represents the fit to a Gaussian distribution (mean \pm s.d., 36.3 ± 7.2 nm). **e**, Step-size histogram for the movement of the deac-aminonucleotide ($n = 267$ steps, 38 MyoV-HMM molecules). The red curve represents the fit to the sum of two Gaussians (shown individually in black lines; mean \pm s.d., 17.5 ± 7.1 nm, 36.0 ± 8.6 nm). **f**, Model for correlation of movement of Alexa-Fluor-568-MyoV-HMM and deac-aminonucleotide binding and dissociation. The red and blue arrows show the position of the centroid of the Alexa-Fluor-568-MyoV-HMM and of the deac-aminonucleotide fluorescence, respectively. See text for description of the model; see also Methods.

(step 2). After deac-aminoATP binds to the nucleotide-free trailing head, this head rapidly dissociates and swings forward to rebind and become the new lead head (steps 3 and 4). Single-molecule and bulk solution studies suggest that the time between detachment of the trailing head, followed by its forward swing and reattachment, is a few milliseconds and is thus much faster than the sampling rate (330 ms) used in our experiments^{7,10,11,19}. Therefore, ATP binding to the trail head, dissociation of that head, and stepping and rebinding are all associated with a 36-nm movement of the MyoV-HMM molecule and a simultaneous 18-nm movement of the deac-amino-nucleotide signal. The binding of deac-aminoATP to the trail head might be expected to produce a transient backward movement of the nucleotide fluorescence centroid, but this is not seen because the trail head quickly detaches and is rapidly moved forward by the power stroke occurring on the lead head.

To confirm the model, lifetimes during the two and one deac-aminonucleotide signal levels were analysed at three different deac-aminoATP concentrations (Fig. 3). We interpret the two to one nucleotide signal decrease to be associated with deac-aminoADP release from the trail head, whereas the one to two nucleotide signal increase is associated with deac-aminoATP binding to that head. Thus, fitting the lifetimes of the high nucleotide signal at 100, 200

and 400 nM deac-aminoATP showed no statistical difference in the rate of deac-aminoADP dissociation (0.82 s^{-1} , 0.79 s^{-1} and 0.90 s^{-1} , respectively; Fig. 3a–c). This is similar to the deac-aminoADP dissociation rate constants measured in solution under identical conditions using stopped-flow spectrofluorimetry (1.2 s^{-1} ; Supplementary Fig. 9c). This would indicate that in our experiments there is no acceleration of the deac-aminoADP release from the trail head and is consistent with stopped-flow kinetic results previously reported¹⁰. An acceleration of the ADP release rate from a positively strained trail head of up to 50-fold was previously predicted if the lead head were to complete its power stroke when both heads were attached⁹. However, an earlier study found that the lead heads were only at the start of their power stroke²⁰, which is consistent with the lack of acceleration of ADP release from the rear head observed in our study. On the other hand, the observed deac-aminoATP binding rates determined by fitting the lifetimes of the low signal level intermediate increased as the deac-aminoATP concentration used was raised from 100 to 200 to 400 nM (0.53 s^{-1} , 0.64 s^{-1} and 1.02 s^{-1} , respectively; Fig. 3d–f). This corresponds to a second order association rate constant of $1.67 \mu\text{M}^{-1} \text{ s}^{-1}$, which is very similar to a value of $2.48 \mu\text{M}^{-1} \text{ s}^{-1}$ measured in solution under identical conditions (Supplementary Fig. 9b).

These results support a model in which the trailing head of the MyoV-HMM molecule releases ADP much more rapidly than the leading head^{2,9,10}. In fact, solution kinetics studies at 20°C demonstrated that the deac-aminoADP dissociation rate (0.48 s^{-1}) from the (presumably) trailing head was 32-times faster than that of the leading head (0.015 s^{-1}) and a similar mechanism occurs with ADP^{10,21}. Inhibition of ADP dissociation from the lead head is thought to be essential for long processive movements. Our results indicate that the main pathway of the MyoV-HMM ATPase is by the central shaded line of intermediates in Fig. 4. The recently detached (formerly rear) head containing ATP or ADP- P_i rapidly swings forward to the leading position where it binds actin (state (1)). On binding to actin, this head quickly releases P_i (state (1) to (2) in Fig. 4)⁷. ADP then

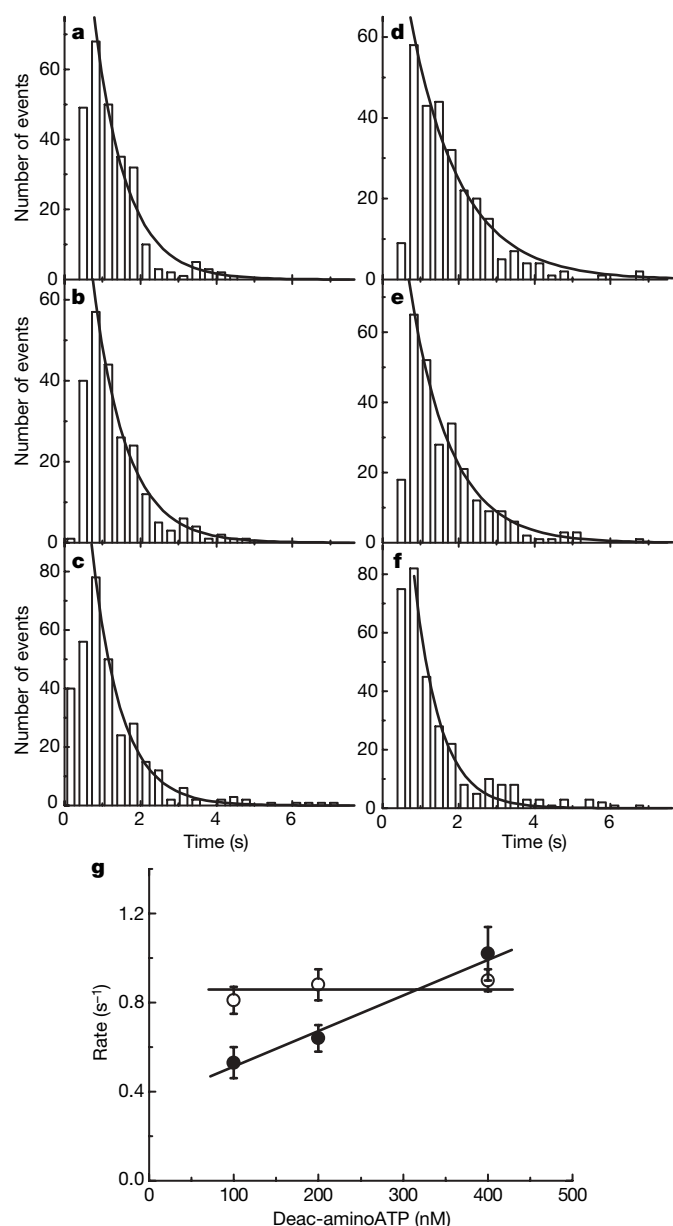


Figure 3 | Histogram of lifetimes of deac-aminonucleotide association and dissociation. **a–c**, Histograms of the lifetimes before deac-aminonucleotide dissociation at 100 nM ($n = 261$ steps, 44 MyoV-HMM molecules), 200 nM ($n = 262$ steps, 38 myosin Va molecule) and 400 nM ($n = 310$ steps, 35 MyoV-HMM molecules) deac-aminoATP. The solid lines represent the exponential fit of the dwell-time distribution. The fitted lifetimes at 100 nM (**a**), 200 nM (**b**), and 400 nM (**c**) deac-aminoATP are $0.85 \pm 0.06 \text{ s}$ ($r^2 = 0.98$), $0.88 \pm 0.07 \text{ s}$ ($r^2 = 0.98$) and $0.77 \pm 0.04 \text{ s}$ ($r^2 = 0.98$; all mean \pm s.d.), respectively, corresponding to rate constants of $0.82 \pm 0.06 \text{ s}^{-1}$, $0.79 \pm 0.07 \text{ s}^{-1}$ and $0.90 \pm 0.05 \text{ s}^{-1}$. **d–f**, Histograms of the lifetimes of deac-aminoATP binding. The fitted lifetimes at 100 nM ($n = 296$; **d**), 200 nM ($n = 267$; **e**) and 400 nM ($n = 309$; **f**) deac-aminoATP are $1.32 \pm 0.17 \text{ s}$ ($r^2 = 0.98$), $1.08 \pm 0.11 \text{ s}$ ($r^2 = 0.98$) and $0.68 \pm 0.08 \text{ s}$ ($r^2 = 0.98$; all mean \pm s.d.), respectively. Note that the number of spots is the same as **a–c**. This corresponds to rate constants of $0.53 \pm 0.07 \text{ s}^{-1}$, $0.64 \pm 0.06 \text{ s}^{-1}$ and $1.02 \pm 0.12 \text{ s}^{-1}$, respectively. Statistical analysis (Student's t -test) between each experimental point showed that the data of ADP dissociation rate from three conditions are not significantly different ($P(T \leq t) = 0.68, 0.09$ and 0.16 of 100 versus 200, 100 versus 400 and 200 versus 400 nM deac-aminoATP, respectively). In contrast, P values of ATP binding rates are significantly different as shown ($P(T \leq t) = 0.01, 0.028$ and 0.02 for 100 versus 200, 100 versus 400 and 200 versus 400 nM deac-aminoATP, respectively). **g**, Concentration dependence of the rates of deac-aminoADP dissociation (open circles) and deac-aminoATP binding (filled circles). The deac-aminoATP binding data were fit by linear regression which gave a slope corresponding to a second order rate constant of $1.67 \mu\text{M}^{-1} \text{ s}^{-1}$. Note the non-zero intercept which is also present in the solution kinetic measurements of deac-aminoATP binding to acto-HMM in Supplementary Fig. 9. A non-zero intercept is predicted by modelling the kinetic mechanism and is seen in some published studies⁵ but is not readily observed because of the large extrapolation from the high nucleotide concentrations typically used in kinetic studies. The horizontal line through the deac-aminoADP dissociation data represents the average of the mean value for the three nucleotide concentrations (0.86 s^{-1}).

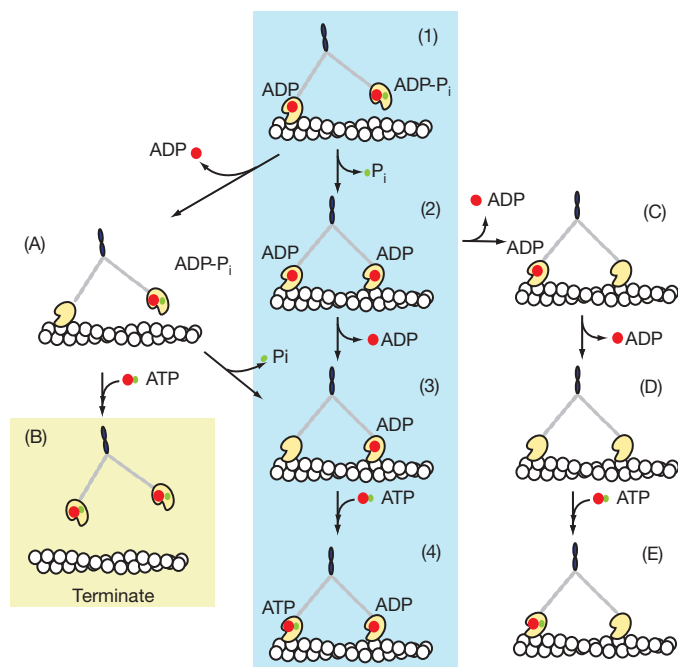


Figure 4 | A scheme of the tight coupling pathway of myosin Va. The mechanism shown by intermediates (1) to (4) is the main pathway of stepping during a MyoV-HMM processive run. Termination of runs would occur by pathway (1) to (A) to (B) or more rarely by (C) to (D) to (E) to (A) to (B). Red spots represent ADP and green spots represent P_i .

dissociates from the trailing head (state (2) to (3)) which allows a new ATP to bind. This results in a rapid detachment of that head, allowing the lead head to undergo its power stroke and repositioning the detached head to become the new lead head (state (3) to (4)). This model accounts for the 36-nm forward steps taken by the Alexa-Fluor-568-MyoV-HMM that occur coincidentally with the 18-nm movement and a doubling of the intensity of the nucleotide fluorescence. An 18-nm backward step of nucleotide fluorescence would occur if deac-aminoADP dissociated first from a lead head (state (2) to (C)). We did not observe such steps, which attest to the high level of strain-dependent gating between the kinetics of the two heads of MyoV-HMM. Termination of runs principally occurs by the route (1) to (A) to (B). This is consistent with most termination cases in which the myosin has only one deac-amino nucleotide bound.

Here we have directly observed the substrate binding and product dissociation steps of single motors moving along their tracks. The data show the relationship between these steps and the mechanism of processive movement of myosin Va on actin. These observations directly demonstrate that, as previously proposed, myosin Va is a tightly coupled motor^{2,4–11,19}. Each step in a processive run involves the binding of an ATP molecule to the trail head of the myosin Va, which is rapidly followed by a 36-nm step along the actin and subsequently by the dissociation of ADP from the trail head. Deac-aminoATP should be a useful analogue for other single-molecule studies such as combined optical trapping and total internal reflection fluorescence (TIRF) microscopy.

METHODS SUMMARY

Protein purification and labelling. Mouse MyoV-HMM, MyoV-S1 and calmodulin were purified as previously described²². Calmodulin was labelled with Alexa Fluor 568 and exchanged for endogenous calmodulins into MyoV-HMM in a similar method to that previously described^{3,22}.

Data acquisition and analysis. The single molecule *in vitro* motility assay was carried out essentially as previously described²². Dual imaging of deac-amino nucleotide and Alexa-Fluor-568-labelled MyoV-HMM were conducted using an Olympus IX81 microscope equipped for two fibre optic input cables using the DualView system²³. Images were taken at a frame rate of 330 ms and the position of each fluorescent spot was determined using the FIONA method¹⁸.

Transient kinetic data. Measurement of the deac-aminoATP binding and deac-aminoADP dissociation were performed on a KinTek stopped-flow spectrofluorimeter as previously described^{10,17}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 March; accepted 3 June 2008.
Published online 30 July 2008.

- Sellers, J. R. & Weisman, L. S. *Myosins: A Superfamily of Molecular Motors, Proteins and Cell Regulation* Vol. 7 (ed. Coluccio, L.) 289–324 (Springer, 2008).
- Vale, R. D. Myosin V motor proteins: marching stepwise towards a mechanism. *J. Cell Biol.* **163**, 445–450 (2003).
- Sakamoto, T., Yildiz, A., Selvin, P. R. & Sellers, J. R. Step-size is determined by neck length in myosin V. *Biochemistry* **44**, 16203–16210 (2005).
- Mehta, A. D. et al. Myosin-V is a processive actin-based motor. *Nature* **400**, 590–593 (1999).
- Rief, M. et al. Myosin-V stepping kinetics: a molecular model for processivity. *Proc. Natl Acad. Sci. USA* **97**, 9482–9486 (2000).
- De La Cruz, E. M., Wells, A. L., Rosenfeld, S. S., Ostap, E. M. & Sweeney, H. L. The kinetic mechanism of myosin V. *Proc. Natl Acad. Sci. USA* **96**, 13726–13731 (1999).
- Rosenfeld, S. S. & Sweeney, H. L. A model of myosin V processivity. *J. Biol. Chem.* **279**, 40110–40111 (2004).
- Purcell, T. J., Sweeney, H. L. & Spudich, J. A. A force-dependent state controls the coordination of processive myosin V. *Proc. Natl Acad. Sci. USA* **102**, 13873–13878 (2005).
- Veigel, C., Schmitz, S., Wang, F. & Sellers, J. R. Load-dependent kinetics of myosin-V can explain its high processivity. *Nature Cell Biol.* **7**, 861–869 (2005).
- Forgacs, E. et al. Kinetics of ADP dissociation from the trail and lead heads of actomyosin V following the power stroke. *J. Biol. Chem.* **283**, 766–773 (2007).
- Veigel, C., Wang, F., Bartoo, M. L., Sellers, J. R. & Molloy, J. E. The gated gait of the processive molecular motor, myosin V. *Nature Cell Biol.* **4**, 59–65 (2001).
- Funatsu, T., Harada, Y., Tokunaga, M., Saito, K. & Yanagida, T. Imaging of single fluorescent molecules and individual ATP turnovers by single myosin molecules in aqueous solution. *Nature* **374**, 555–559 (1995).
- Ishijima, A. et al. Multiple- and single-molecule analysis of the actomyosin motor by nanometer piconewton manipulation with a microneedle: Unitary steps and forces. *Biophys. J.* **70**, 383–400 (1996).
- Tokunaga, M., Kitamura, K., Saito, K., Iwane, A. H. & Yanagida, T. Single molecule imaging of fluorophores and enzymatic reactions achieved by objective-type total internal reflection fluorescence microscopy. *Biochem. Biophys. Res. Commun.* **235**, 47–53 (1997).
- Oiwa, K. et al. Comparative single-molecule and ensemble myosin enzymology: sulfoindocyanine ATP and ADP derivatives. *Biophys. J.* **78**, 3048–3071 (2000).
- Webb, M. R., Reid, G. P., Munasinghe, V. R. & Corrie, J. E. A series of related nucleotide analogues that aids optimization of fluorescence signals in probing the mechanism of P-loop ATPases, such as actomyosin. *Biochemistry* **43**, 14463–14471 (2004).
- Forgacs, E. et al. Kinetic mechanism of myosinV-S1 using a new fluorescent ATP analogue. *Biochemistry* **45**, 13035–13045 (2006).
- Yildiz, A. et al. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300**, 2061–2065 (2003).
- Dunn, A. R. & Spudich, J. A. Dynamics of the unbound head during myosin V processive translocation. *Nature Struct. Mol. Biol.* **14**, 246–248 (2007).
- Burgess, S. et al. The prepower stroke conformation of myosin V. *J. Cell Biol.* **159**, 983–991 (2002).
- Rosenfeld, S. S., Houdusse, A. & Sweeney, H. L. Magnesium regulates ADP dissociation from myosin V. *J. Biol. Chem.* **280**, 6072–6079 (2005).
- Sakamoto, T. et al. Neck length and processivity of myosin V. *J. Biol. Chem.* **278**, 29201–29207 (2003).
- Thirumurugan, K., Sakamoto, T., Hammer, J. A. III, Sellers, J. R. & Knight, P. J. The cargo-binding domain regulates structure and activity of myosin 5. *Nature* **442**, 212–215 (2006).
- Wang, F. et al. Effect of ADP and ionic strength on the kinetic and motile properties of recombinant mouse myosin V. *J. Biol. Chem.* **275**, 4329–4335 (2000).
- Webb, M. R. & Corrie, J. E. Fluorescent coumarin-labeled nucleotides to measure ADP release from actomyosin. *Biophys. J.* **81**, 1562–1569 (2001).
- Sakamoto, T., Amitani, I., Yokota, E. & Ando, T. Direct observation of processive movement by individual myosin V molecules. *Biochem. Biophys. Res. Commun.* **272**, 586–590 (2000).
- Thompson, R. E., Larson, D. R. & Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775–2783 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Zhang, A. Smith and G. Reid for technical assistance; E. Yokoi for the TIRF illuminator to combine two optical cables; C. Fanghella for technical help in the calculation of the number of photons; and C. Weaver for help with Metamorph. We are appreciative of the critical comments on the manuscript made by P. J. Knight and E. Homsher. M.R.W. was supported by

the Medical Research Council, UK. J.R.S. and T.S. were supported by the National Heart, Lung and Blood Intramural Program. H.D.W. and E.F. were supported by NIH EB00209 and a postdoctoral fellowship from the American Heart Association.

Author Contributions Single molecule motility experiments and data analysis were performed by T.S. and kinetic experiments data by E.F. Deac-aminoATP/ADP were provided by M.R.W. T.S., J.R.S. and H.D.W. participated in the conception of the

experiment. T.S. wrote the first draft of the manuscript and all authors participated in producing the final version. All authors participated in discussion and interpretation of the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.R.S. (Sellersj@mail.nih.gov).

METHODS

Preparation of proteins. Mouse MyoV-HMM and MyoV-S1 were purified from Sf9 cells after infection with baculoviruses driving the expression of the HMM (or S1) and calmodulin²⁴. Calmodulin was purified from bovine testes and labelled with Alexa Fluor 568 succinimidyl ester (Invitrogen)²². The molar ratio of Alexa Fluor 568 per calmodulin was determined to be 1.8 from the absorbance in solution and a molar extinction coefficient of $91,300 \text{ M}^{-1} \text{ cm}^{-1}$ for Alexa Fluor 568 and $\epsilon^{0.1\%} = 0.18$ at 280 nm for calmodulin. The labelled calmodulin (molar ratio of 20 per MyoV-HMM) was exchanged with endogenous calmodulin as previously described²². This resulted in an average of six Alexa Fluor 568 dyes per MyoV-HMM, as determined spectrophotometrically. This value was confirmed by comparing the intensity of single Alexa Fluor 568 molecules bound nonspecifically to a surface (2,000 a.u.) with that of the average intensity of Alexa-Fluor-568-MyoV-HMM (12,000 a.u.). Biotinylated actin and biotinylated BSA were prepared²² and deac-aminoadp and deac-aminoadp were synthesized as described previously²⁵.

Emission and excitation spectra of $0.5 \mu\text{M}$ deac-aminoadp in the presence and absence of $1 \mu\text{M}$ MyoV-HMM were taken with a Fluoromax-3 spectrofluorimeter (HORIBA Jobin Yvon, NJ) using 2 nm slits.

Two-line total internal reflection fluorescence microscopy. Alexa-Fluor-568-labelled MyoV-HMM and deac-aminoadp were observed by objective-type TIRF microscopy using an Olympus IX81 microscope and a $\times 60$, 1.45 numerical aperture PlanApo objective lens with two magnifying (relay) lenses ($\times 1.6$ in the microscope and $\times 2.5$ in front of the camera). The temperature was kept at 25°C with an environmental box (Precision plastics). To visualize two colours of fluorescence simultaneously, we used the 568 nm line from an Ar-Kr Laser (model I70C, Spectra physics) for Alexa Fluor 568 and the 442 nm line from a He-Cd Laser (model IK41711-G, KIMMON) for deac-aminoadp. Both laser lines were combined by an acousto-optical tunable filter (Prairie Technologies), which also controlled the laser power. After the acousto-optical tunable filter, the two laser lines were separated by a dichroic mirror onto optical fibres. This allows both wavelengths to be in focus at the same time. The two fibres are guided to individual TIRF illuminators located at the rear end of the microscope. Illumination at 442 nm was by the Olympus TIRF apparatus and the illumination at 568 nm was by the position usually occupied by the mercury arc lamp housing. The two laser lines from the two illuminators were combined with a dichroic mirror and introduced into the objective lens. The power of both the 442 nm and 568 nm beams was 10 mW to 20 mW in front of the objective lenses. The emitted light was passed through a dual line dichroic mirror (442/568, Chroma) and split by a dichroic mirror (552dcr, Chroma) in the Dual-View system (Optical Insights). Fluorescence was detected by an EMCCD camera (DV897, 512BV, Andor technology), at -90°C with a gain of either 400 or 1,000. Images were digitized by using Metamorph (MSD/Molecular Device ver.7.1).

Intensity measurement of deac-aminoadp bound either to the surface or to MyoV-HMM. To test whether the fluorescence of deac-aminoadp increased in intensity upon binding to MyoV-HMM in the microscope, we directly observed single molecules of deac-aminoadp in the presence and absence of MyoV-HMM.

First, 10 pM deac-aminoadp was added into a flowcell coated with 0.1% nitrocellulose and incubated for 2 min at room temperature. Free deac-aminoadp was washed out using motility assay buffer (40 mM KCl , 20 mM MOPS , 4 mM MgCl_2 , 0.1 mM EGTA , $1 \mu\text{M}$ calmodulin, and 50 mM DTT , pH 7.5, 25°C). The solutions also included an oxygen scavenging system composed of $25 \mu\text{g ml}^{-1}$ glucose oxidase, $45 \mu\text{g ml}^{-1}$ catalase and 2.5 mg ml^{-1} glucose. Deac-aminoadp was imaged at 442 nm using the TIRF microscopy set up described above at EMCCD camera gains of 1,000 (Fig. 1a, e) and 400 (Fig. 1b, f); images were acquired in 330 ms windows. On a second slide, 10 pM MyoV-HMM was added into the flowcell and incubated for 2 min at room temperature. Free MyoV-HMM was washed out using motility assay buffer. Deac-aminoadp (10 nM) was added into the flowcell and the sample was imaged at 442 nm and 568 nm at the same two camera gains. Under these conditions at the single-molecule level in the microscope, the intensity of deac-aminoadp increased 4 fold upon binding to MyoV-HMM compared to the 25 fold change in solution.

Single-molecule motility assay and data analysis. Single-molecule motility assays were performed as previously described²⁶. Position data for Alexa-Fluor-568-MyoV-HMM and deac-aminoadp were analysed by FIONA¹⁸. The integrated intensities of 15×15 pixel areas were measured at the indicated concentrations using Metamorph. To observe single-molecule movements of MyoV-HMM and deac-aminoadp simultaneously, we changed the concentration of Alexa-Fluor-568-MyoV-HMM to reduce background. At 100 nM and 200 nM deac-aminoadp concentrations, 200 pM Alexa-Fluor-568-MyoV-HMM was used, whereas at 400 nM deac-aminoadp, 4 pM Alexa-Fluor-568-MyoV-HMM was used. Steps were identified by eye and marked by hand.

Run lengths of Alexa-Fluor-568-MyoV-HMM were measured with either 1 mM ATP or 1 mM deac-aminoadp. Actin filaments were labelled with 10% Alexa-Fluor-488-phalloidin and 90% phalloidin. The determination of run length was performed as described previously²², except that only myosin molecules that dissociated before reaching the end of actin filaments were scored. The average length of an actin filament was $12.5 \mu\text{m}$. Velocities with various deac-aminoadp concentrations were measured by time lapse in which data were taken at 10 s intervals with a 300 ms exposure time. Sequential images were taken to analyse velocity.

Determination of number of photons. We determined the number of photons from the integrated intensity of a 10×10 pixel image of each chosen spot. Deac-aminoadp was bound nonspecifically to the surface or to MyoV-HMM, which was bound on a nitrocellulose-coated surface. The estimated total number of photons in the spot at various camera gains was calculated as previously described²⁷. Alternatively, the number of photons, n , was calculated using an equation $n = 60.14 \cdot I/a$ provided by Andor Technology (data not shown) in which 60.14 is the camera sensitivity at 10 MHz, electron multiplying amplifier, $\times 1.0$ preamp setting (electron per A/D count), a is the percentage of quantum efficiency of the camera at the appropriate wavelength, and I is the detected integrated intensity. Results obtained by the two methods gave reasonable agreement. At least 10,000 photons are required to obtain 2.5 nm localization¹⁹. Determination of localization accuracy from single-molecule fluorophores has been calculated by theoretical equations²⁷ and measured experimentally¹⁸.

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Suppose you're an assistant professor, striving for tenure at your university. You're diligently conducting your research, scrambling to put together your publications and tending to your teaching duties. Then one day your dean phones you up with an ominous request: "I'd like to talk to you about the negative ratings you've received on RateMyProfessors.com."

In the United States, websites such as RateMyProfessors.com, which allow students to post anonymous reviews of their university professors, are growing in popularity. The comments range from the positive ("The definition of a perfect professor"), to the disparaging ("He is horrible"), to the slightly tawdry ("Quite possibly the hottest prof you'll ever find"). Perhaps unsurprisingly, these sites are proving less of a hit with US professors, many of whom are uneasy about the unvetted comments that are allowed.

But in early August, an analysis of the effectiveness of RateMyProfessors.com had some cautious praise for the site and its approach (J. Otto, D. A. Sanford and D. N. Ross *Assess. Eval. High. Educ.* **33**, 355–368; 2008). The authors found that the feedback and ratings seemed surprisingly free of the universally high or low ratings that might be expected, given that the site is likely to attract students predisposed to wanting to praise or damn their professors. The site, suggest the authors, could potentially be seen as a valid measure of teaching effectiveness. Although the possibility of individual bias remains, they note that if further research backs their initial assessment, sites such as RateMyProfessors.com could be used to help inform decisions on hiring and promoting faculty members.

That would be a radical shift — although it is hard to contest the idea that honest feedback is a valid metric for teaching performance. But for the ratings sites to gain true credibility, some degree of policing is needed. The raters' background and study course should be revealed, and only students who actually attended the relevant class should be allowed to post a comment on it. As for the relevance of how 'hot' a professor is, this will just have to be left up to the individual discerning student.

Gene Russo is editor of Naturejobs.

CONTACTS

Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:

Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW: Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Dennis Chu

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718

Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Suppose you're an assistant professor, striving for tenure at your university. You're diligently conducting your research, scrambling to put together your publications and tending to your teaching duties. Then one day your dean phones you up with an ominous request: "I'd like to talk to you about the negative ratings you've received on RateMyProfessors.com."

In the United States, websites such as RateMyProfessors.com, which allow students to post anonymous reviews of their university professors, are growing in popularity. The comments range from the positive ("The definition of a perfect professor"), to the disparaging ("He is horrible"), to the slightly tawdry ("Quite possibly the hottest prof you'll ever find"). Perhaps unsurprisingly, these sites are proving less of a hit with US professors, many of whom are uneasy about the unvetted comments that are allowed.

But in early August, an analysis of the effectiveness of RateMyProfessors.com had some cautious praise for the site and its approach (J. Otto, D. A. Sanford and D. N. Ross *Assess. Eval. High. Educ.* **33**, 355–368; 2008). The authors found that the feedback and ratings seemed surprisingly free of the universally high or low ratings that might be expected, given that the site is likely to attract students predisposed to wanting to praise or damn their professors. The site, suggest the authors, could potentially be seen as a valid measure of teaching effectiveness. Although the possibility of individual bias remains, they note that if further research backs their initial assessment, sites such as RateMyProfessors.com could be used to help inform decisions on hiring and promoting faculty members.

That would be a radical shift — although it is hard to contest the idea that honest feedback is a valid metric for teaching performance. But for the ratings sites to gain true credibility, some degree of policing is needed. The raters' background and study course should be revealed, and only students who actually attended the relevant class should be allowed to post a comment on it. As for the relevance of how 'hot' a professor is, this will just have to be left up to the individual discerning student.

Gene Russo is editor of Naturejobs.

CONTACTS

Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
Natureevents:

Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW: Nils Moeller (4953)

Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Dennis Chu

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718

Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com

MOVERS

Edward Seidel, director, Office of Cyberinfrastructure, National Science Foundation, Arlington, Virginia



2003-08: Director, Center for Computation and Technology, and professor of floating point systems in the departments of physics and astronomy, and computer science, Louisiana State University, Baton Rouge

1996-2003: Professor of numerical relativity, Max Planck Institute for Gravitational Physics, Potsdam, Germany

Edward Seidel has been fascinated by space exploration since he was an 8-year-old *Star Trek* fan, a passion that he retained as a young man. But as time went by, his attention turned from space travel to computational astrophysics.

He began his career studying mathematics and physics at the College of William and Mary in Williamsburg, Virginia, before going on to earn a PhD in relativistic astrophysics at Yale University in New Haven, Connecticut. Seidel followed this up with postdoctoral positions at Washington University in St Louis and the University of Illinois at Urbana-Champaign, before computer guru Larry Smarr hired him as a research scientist for the latter institution's National Center for Supercomputing Applications. The move turned out to be one of the most important of Seidel's career. Smarr was at the vanguard of computational astrophysics research, and showed Seidel the importance of supercomputing networks.

After seven years in Illinois, Seidel moved to Germany to help set up the Max Planck Institute for Gravitational Physics, which was founded in 1995 as part of the eastward expansion of the Max Planck Society started after Germany's reunification in 1990. Former colleague Bernard Schutz, now head of the institute's department of astrophysical relativity, lauds not only Seidel's work on black holes, but also his help in advancing communications among European astrophysicists as co-founder of the EU Astrophysics Network.

In 2003, Seidel returned to the United States to become director of the newly established Center for Computation and Technology at Louisiana State University. There, he advanced the use of vast computer networks for studies of complex natural phenomena such as black holes. But what he enjoyed most was the interdisciplinary research, seeing the potential of cyberinfrastructure to reach far beyond astrophysics.

In Seidel's new position at the National Science Foundation's Office of Cyberinfrastructure, he will deal with similar issues but on a more international scale. He will be responsible for dispensing money to scientists for computer facilities. "I will get the chance to work with investigators from many scientific fields to develop cyberinfrastructure," says Seidel. "The position requires the ability to listen carefully to everyone in the community." This seems like the perfect vocation for Seidel, says Schutz, who adds: "He respects everyone, from principal investigators right down to the youngest graduate student."

Maria Rossbauer

NETWORKS & SUPPORT

Boosting Brazilian bioenergy

The aim of Brazil's US\$46-million bioenergy research programme (BIOEN) is to keep the country at the cutting edge of biofuels research and development — in part by attracting bright young minds.

After the United States, Brazil is the world's largest ethanol producer. Maintaining its position as a biofuels leader will require improved biofuel-processing techniques, says Carlos Henrique de Brito Cruz, the scientific director of the State of São Paulo Research Foundation (FAPESP), which runs BIOEN. "We need to build a critical mass of top scientists in the fields of plant physiology, bioinformatics and enzymatic hydrolysis to achieve this goal," he says.

BIOEN's projects aim to foster an interdisciplinary approach that enhances biofuels processing at every stage — from plants' photosynthesis to the enzymatic fermentation of sugar cane to create ethanol. The programme will also focus on the social impacts of biofuels production — such as unintended effects on agricultural markets — says Glauca Souza, BIOEN's biomass programme coordinator.

The funds — from FAPESP, Brazil's National Council for Scientific and Technological Development, the State of Minas Gerais Research Foundation,

and Dedini, one of the private companies involved — will promote cooperation between academia and industry.

The Young Investigator Award is the cornerstone of BIOEN, and will fund about 20 scientists' first independent research programmes. Each will receive at least \$200,000 for projects lasting up to four years, including an annual salary of \$39,000. The monies are intended to help the young researchers to establish laboratories in Brazil — an achievement that will enhance their future employment opportunities.

BIOEN's partner companies are hiring as well, augmenting São Paulo's career opportunities. Bioenergy equipment manufacturer Dedini supports university-based research projects, and is hiring senior researchers and chemical engineers with higher degrees in energy science to help produce ethanol from cellulose. "With BIOEN, we hope to continue improving our hydrolysis efforts to reach the commercial scale," says José Olivério, Dedini's vice-president of research and development.

BIOEN is expected to gain \$130 million of investment during the next five years, which should mean additional opportunities for engineers and scientists.

Virginia Gewin

POSTDOC JOURNAL

Diagnosing mysteries

For the first time in three years, something stopped me thinking about my research. The other week, I woke up with a severe headache and numbness on the right side of my face. These symptoms sent me to the emergency room, and then to a team of neurologists, whose care I have been under. After a slew of tests, including a clean MRI scan, there was no precise diagnosis of my ailment. I was sent home with pain medication and a recommendation for a facial X-ray, a bone scan and a spinal tap. In other words, the doctors are now shooting in the dark, with little idea as to the source of my illness.

On reflection, my thoughts returned to my own research. Although biologists and medical doctors have made great strides in resolving some of life's mysteries, there is so much more we do not yet know. Lying in the hospital bed sipping cold chicken soup, I realized that I have less control over my research direction than I once believed. In truth, as I explore the mechanistic nuances of plant growth, I simply expose more of the unexplored, which often leads me to shoot in the dark to get to the next step.

My intuition as to where I should aim can only take me so far and, ultimately, like my doctors, I am often left scratching my head in wonder and amazement.

Zachary Lippman is a postdoctoral fellow at the Hebrew University of Jerusalem's faculty of agriculture.

Gigatech

Testing, testing ...

David Langford

Seven minutes ...

You never get the future you expect. We all knew, or thought we knew, that the next big thing would be a very small thing indeed: nanotechnology, molecular assemblers, microscopic robots unclogging arteries, restoring synapses lost to Alzheimer's, and generally clearing the way to immortality.

Instead we have a massive lump of gigatechnology. We're lumbered with the Orrery — the ultimate geek gift. And as project leader at Orbital HQ, mine is the first head to roll if the test run fails.

Six minutes.

There's not much precedent for the appearance in high Earth orbit of an object 978 kilometres in diameter and glowing pale blue. Once everyone had recovered from natural panic and announcements of the End Times, the thing up there had to be investigated. While the US shuttle programme was being rejigged, though, there came a slight disruption as a sequence of massive electromagnetic pulses knocked out two-thirds of our satellite communications. The Orrery had broadcast its instruction manual.

"I still can't believe it," said Carson in the command pod. "We're going to move a bloody asteroid!"

Five minutes.

As a name, 'Orrery' was a lucky guess suggested by its three-dimensional spider-web construction of slender curved rods with vari-sized nests of globes at their countless (actually my team has counted 5,271,009) intersection points. An orrery, lower-case, is a crude rod-and-ball model of the Solar System. The Orrery is a representation of our local galactic region — relating to astronomy rather as the London Underground's iconic 'diagram of lines' map is distantly connected to city geography.

A galactic transport map? An under-space wormhole network across the stars? "My God, it's full of worms!" Another inspired guess, but wrong.

Four minutes.

It was reassuring when the first shuttle crew confirmed that the Orrery's gentle glow is confined to the visible spectrum. It was less comforting to learn that the still unidentified construction material is impervious to neutrinos. As the Joint Physics Advisory Committee concluded: "How the hell did they do that?" Even the

artefact's orbital mechanics are subtly too good to be true. Without visible course correction, it sweeps out a perfect circle that should before long be detectably perturbed, but isn't.

Conundrums like this are a useful distraction from gnawing thoughts about the worst-case scenario in just ...

Three minutes.

The Orrery's electromagnetic shriek was surprisingly easy to decipher. Clearly this was intentional. Ingenious fractal encoding caused the shape of the message to be implicit in the shape of its envelope. There are ambiguities, but the manual conveys that life is to be valued; that the Orrery is a tool to preserve life; that it laughs at assumed limits like speed-of-light; and that it should be operated as follows to adjust matters in its area of effect.

We're going to move a bloody asteroid.

Two minutes.

Not a transport map but an interactive map, intimately linked with the territory. A telefactoring device. The manual's eye-opening example is the hypothetical case of impending comet impact on one's planet. Tweak the Orrery, and the successor to the Dinosaur Killer can be flipped into a new trajectory almost as soon as it's detected.

So we're gaily pretending Ceres is an Earth-grazer, and making a small trial adjustment ...

One minute.

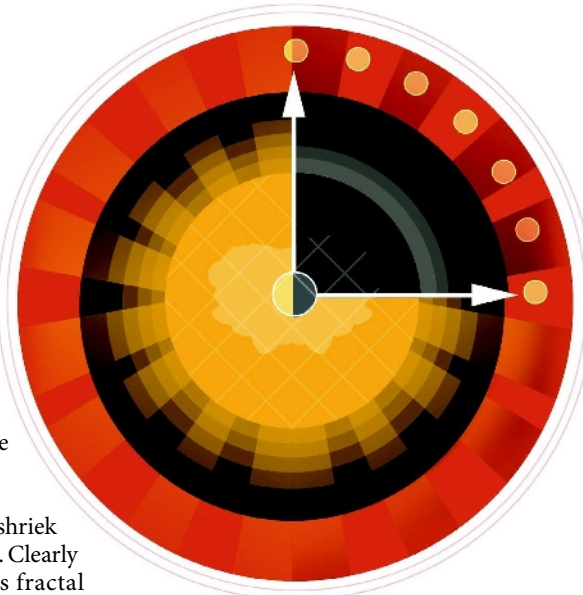
There are safeguards, thank God. You wouldn't want some casual impact from space junk to reverse the Sun's rotation or set Jupiter on fire. But this isn't what the science-fiction fans call a Big Dumb Object. Like a child-proof medicine bottle, the Orrery resists random prods and twists. It understands a pattern of shaped atomic charges at the appropriate node.

Of course the thing has to be tried. How could we not?

Thirty seconds.

Orbital HQ was quiet apart from the faint whisper of air recyclers. The Orrery hung out there, glowing tranquilly. Our triple nuclear package was invisible at this distance. I didn't like to break the silence but wished Carson or somebody would.

What's in it for the builders of gigatechnology? Intelligent self-interest, according to the boffins, who are talking about sentience on a galactic scale. The gigaminds are emergent phenomena of stellar arrangement, maybe with gravitons for neurons.



These builders think big because they can think no other way, and the Orrery is their tiniest possible instrument of precision. With this device, we rude mechanicals can regulate the layout of stars in our local galactic zone.

Twenty seconds.

Our mission, should we choose to accept it, will be to toil away like molecular assemblers, microscopic handlers unclogging spatial arteries, restoring stellar synapses lost to the gigascale equivalent of Alzheimer's ... In short: you never get the future you expect. We have met the nanotechnology, and it is us.

Ten seconds.

My script had something about another giant leap for mankind, but I couldn't bring myself to say that crap. I muttered: "Here goes."

Zero.

Deep amid the fractal tangles of the Orrery, a dazzling fireball bloomed in vacuum. A bad one. Maybe the naked eye can't distinguish a triple blast from a single one, but our readouts certainly could. What a dozen trial runs tell you isn't necessarily true. Charge number 3 had blown milliseconds early and zapped its companions with friendly fire.

"Fratricide incident," I recited uselessly.

Carson, more to the point, said: "Bugger."

Heads will roll for this. Or maybe they won't, because now I have that worst-case scenario to panic about. Earth still in place, check. Moon ditto. But there's 8.3 minutes of light stacked up on its way from a Sun that may no longer be there, or may be horribly changed. One minute already gone. A new countdown is under way, the longest ever:

Seven minutes ...

David Langford, a former physicist turned SF author and critic, lives in a Reading house whose mantelpieces are crowded with 28 Hugo Awards.